Beyond Scylla and Charybdis:
Four Essays on Latent Heterogeneity in Economic Behavior

Dissertation
for the Faculty of Economics, Business Administration and Information Technology of the University of Zurich

to achieve the title of
Doctor of Economics

presented by
Adrian Bruhin
from Zurich

approved at the request of

Prof. Dr. Rainer Winkelmann
Prof. Dr. Ernst Fehr

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich hereby authorizes the printing of this Doctoral Thesis, without thereby giving any opinion on the views contained therein.

Zurich, October 22, 2008

The Dean: Prof. Dr. Dr. Josef Falkinger

To my parents Monika and Rolf Bruhin.

Acknowledgments

While working on this thesis I enjoyed the ongoing support and inspiration of various colleagues and friends.

First of all, I like to thank my two supervisors Rainer Winkelmann and Ernst Fehr. Rainer Winkelmann not only introduced me to, but also formed my basic views on empirical economic research. His constant encouragement, support and helpful critique made this endeavor possible in the first place. Furthermore, I greatly appreciate the highly productive and liberal atmosphere he maintains at his chair where I enjoyed working as one of his research assistants. Many thanks also go to my co-supervisor, Ernst Fehr, who supported this project through his unparalleled experience and helped me develop my own research strategy.

I owe my best thanks to my friends and co-authors, Helga Fehr and Thomas Epper. They decisively contributed to this thesis with their specific know-how and their nearly boundless efforts. Thomas Epper greatly helped to design the experiments and Helga Fehr decisively supported this project through her deep and profound knowledge about the literature on decision making under risk. Our interactions and discussions, both professional and private, made daily work a very inspiring and enjoyable experience.

Finally, my sincere thanks go to Stefan Boes, Kevin Staub, Michael Wolf, Heini Murer, David Moser and Sandra Hanslin who all supported me in many different ways during the past two years.

Contents

Chapter 1

Introduction

1.1 Finite Mixture Models

Most economic models are limited to analyzing the behavior of a representative agent and, consequently, make the implicit assumption that either individuals are homogeneous or that individual heterogeneity does not matter for the aggregate outcome. However, in their seminal papers, Haltiwanger and Waldman assume a mixture of "sophisticated" and "naive" types and show, that if markets are imperfect and there is strategic complementarity, i.e. an increase in one agent's action causes an incentive for the other agents to increase their actions too, minorities can have a decisive influence on the market's outcome (Haltiwanger and Waldman, 1985, 1989). Recent empirical evidence in experimental economics indicates that under strategic complementarity a minority of irrational agents may, indeed, drive the market's outcome (Fehr and Tyran, 2005). Thus, to avoid potential aggregation bias, researchers in empirical economics should take individual heterogeneity into account, which results in the following trade-off: On the one hand, a representative agent approach, which is parsimonious and easy to interpret, completely neglects heterogeneity, but on the other hand, estimating economic behavior at the individual level, which requires a lot of parameters and results in a plethora of estimates, may demand too much from the data. To escape this dilemma, empirical economists may apply finite mixture models, which offer a compromise between completely ignoring individual heterogeneity and running into difficulties when estimating individual by individual.

In contrast to common formulations in econometrics, finite mixture models relax the assumption of homogeneity and assume the population as being made up of a finite number of C different groups. Each of these groups has its own data generating process with component density $g_c(\mathcal{X}; \theta_c)$, where $\mathcal{X}$ represents the data and θ_c denotes a group-specific vector of parameters. The component densities can principally be chosen among all valid density functions and need not necessarily belong to the same distributional family. This renders finite mixture models an extremely flexible tool, either for fitting heterogeneous data sets with several distinctly different types or, from a more data-oriented perspective, for semi-parametrically approximating any complex empirical density by a mix of Gaussian components (Priebe, 1994).

As noted by McLachlan and Peel (2000), the history of finite mixture models dates back over a century to Pearson (1894) who fitted a mixture of two normally distributed components to identify two subpopulations of crabs differing in their average size. However, in spite of their potential flexibility, their complex structure prevented finite mixture models from becoming popular in applied empirical research for over 80 years. Not until the advent of the EM algorithm (Dempster et al., 1977), combined with the rapidly declining price of computing power, have applications of finite mixture models surged in various fields, such as biology, medicine, computer and social science. In political science, for example, McCutcheon (1987), contributed an early survey of latent class analysis, an application of finite mixture modelling to contingency tables.

In econometrics, Heckman and Singer (1984) aimed at consistently estimating structural parameters in the presence of unobserved heterogeneity. In contrast to classical random coefficient models, which rely on distributional assumptions, their approach uses a finite mixture model for semi-parametrically approximating the (continuous) distribution of unobservables and, therefore, does not intend to classify individuals according to afinite number of C distinct types. In his seminal paper, however, Hamilton (1989) introduced finite mixture models to time series analysis where they are usually referred to as regime switching models and used to identify "[...] occasional, discrete shifts" in the analyzed series, an approach which comes close to the idea of characterizing different groups.

The concept of applying finite mixture regression models to identify distinctly different behavioral types and to classify each individual stochastically as one of these types, which is central to all four applications in this thesis, has emerged every now and then in experimental economics since the nineties of the past century (see Stahl and Wilson (1995), El-Gamal and Grether (1995), and Houser et al. (2004) for example). As they allow controlling for different behavioral types while being relatively parsimonious, finite mixture models offer a compromise in the mentioned trade-off between staying completely agnostic about individual heterogeneity by estimating a representative agent's behavior and demanding too much from the data by running the estimation at the individual level. In the context of endogenously classifying individuals to a finite number of distinctly different behavioral types, finite mixture regression models have proven to be valuable tools to control for unobserved heterogeneity in experimental data sets and, recently, have even been applied in a similar manner for analyzing survey data (Clark et al., 2005).

1.1.1 General Structure

The characteristic structure of a finite mixture model arises quite naturally: As it is *a priori* unknown which individual $i \in \{1, \ldots, N\}$ is associated with which group, the researcher faces latent heterogeneity and has to write the individual contribution to the finite mixture model's density as the sum of the component densities $g_c(\mathcal{X}_i; \theta_c)$ weighted by the probabilities π_c that individual i belongs to group c:

$$f(\mathcal{X}_i; \theta) = \sum_{c=1}^{C} \pi_c \, g_c(\mathcal{X}_i; \theta_c) \,. \tag{1.1}$$

These probabilities π_c, which have to sum up to one, represent $C-1$ additional parameters of the model and are equal to the respective groups' relative sizes, since all observations are drawn with the same probability.

To fit his model, Pearson (1894) applied the method of moments which involved the tremendous computational effort of finding the roots of a ninth-degree polynomial. However, their complex likelihood function, the cost for their flexibility, prevented finite mixture models from becoming popular in applied empirical research even after the advent of the more efficient method of maximum likelihood and inexpensive personal computers. Indeed, it is widely known that even after taking logs, a finite mixture model's likelihood function

$$\ln L(\Psi; \mathcal{X}) = \sum_{i=1}^{N} \ln \sum_{c=1}^{C} \pi_c \, g_c(\mathcal{X}_i; \theta_c) \,, \tag{1.2}$$

where $\Psi = (\theta_1', \ldots, \theta_C', \pi_1, \ldots, \pi_{C-1})'$, is highly non-linear, potentially multimodal and can even be unbounded, which makes it very hard to be estimated by standard optimization routines implemented in today's statistical software packages.

1.1.2 Estimation

Things have changed with the development of the Expectation Maximization (EM) algorithm by Dempster et al. (1977) who focus on the issue of latent heterogeneity and view

$\mathcal{X}$ as an incomplete data set. They note that if the data were complete, i.e. individual group membership were observed and indicated by $t_{ic} \in \{0,1\}$, the individual density would be given by

$$\tilde{f}\left(\mathcal{X}_i, t_i; \theta\right) = \prod_{c=1}^{C} \left[\pi_c \, g_c\left(\mathcal{X}_i; \theta_c\right)\right]^{t_{ic}} . \tag{1.3}$$

Since equation (1.3) would involve a product over all components, and not a sum such as equation (1.1) does, the complete-data log likelihood function would simplify to

$$\ln \tilde{L}\left(\Psi; \mathcal{X}, t\right) = \sum_{i=1}^{N} \sum_{c=1}^{C} t_{ic} \left[\ln \pi_c + \ln g_c\left(\mathcal{X}_i; \theta_c\right)\right] , \tag{1.4}$$

where the maximum likelihood estimates of the relative group sizes, $\hat{\pi}_c = 1/N \sum_{i=1}^{N} t_{ic}$, would be given analytically and could be obtained separately from the estimates of the group-specific parameters, $\hat{\theta}_c$. The EM algorithm operates on the complete-data log likelihood by proceeding iteratively in two steps, E and M. During the E-step of the $(k+1)$-th iteration the *a posteriori* probabilities of individual group membership are computed according to Bayes' law given the actual fit to the data, $\Psi^{(k)}$:

$$\tau_{ic}^{(k+1)}\left(\mathcal{X}; \Psi^{(k)}\right) = \frac{\pi_c^{(k)} \, g_c\left(\mathcal{X}_i; \theta_c^{(k)}\right)}{\sum_{m=1}^{C} \pi_m^{(k)} \, g_m\left(\mathcal{X}_i; \theta_m^{(k)}\right)} . \tag{1.5}$$

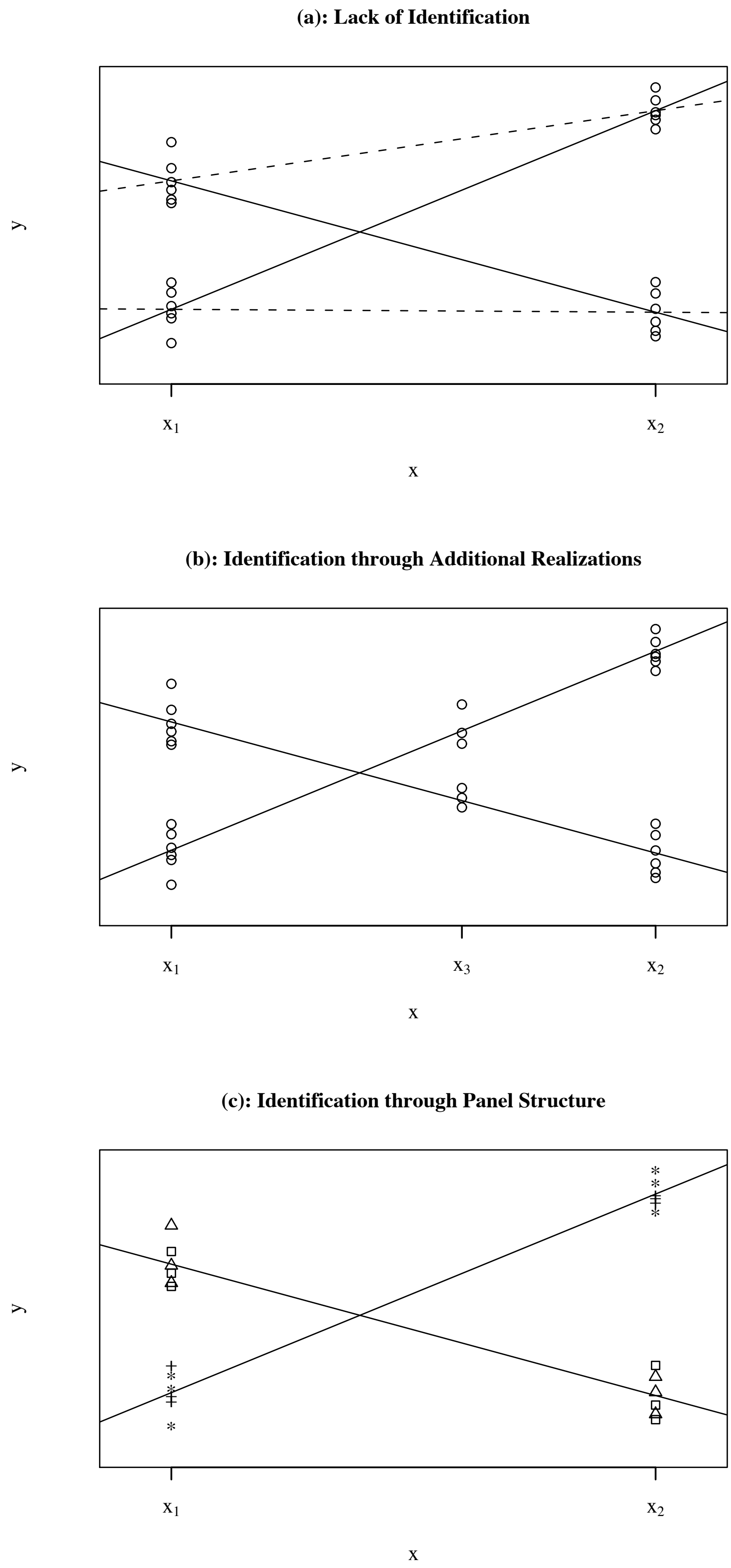
In the following M-step, with these $\tau_{ic}^{(k+1)}\left(\mathcal{X}; \Psi^{(k)}\right)$ replacing the unobserved indicators of individual group membership, t_{ic}, the complete-data log likelihood (1.4) is maximized, providing an update of the parameter vector $\Psi^{(k+1)}$. Note that these *a posteriori* probabilities of individual group membership are not only used in the M-step, but they also provide a tool for assigning each individual in the sample to one of the C groups. Thus, finite mixture models may serve as statistically well grounded tools for endogenous individual classification. Dempster et al. (1977) showed that the likelihood never decreases from one iteration to the next and that the EM algorithm, as any hill climbing algorithm, converges monotonically to the maximum which is closest to its initial values, $\Psi^{(0)}$.

So, even if the EM algorithm helps coping with the characteristic non-linearity of a finite mixture model's likelihood it may still be plagued by its potential multimodality. To overcome such a tendency of converging towards local maxima in complex finite mixture models, Celeux et al. (1996) suggest using stochastic extensions of the EM algorithm. In each iteration, these stochastic versions of the EM algorithm all have a non-zero probability of leaving a once taken path to convergence and continuing in a different region of the likelihood function. Such a procedure results in an improved robustness to the objective function's potential multimodality but comes at the cost of even higher computational demands. However, a hybrid algorithm, as suggested by Render and Walker (1984) in their treatise on maximum likelihood estimation of mixture densities, which first uses an EM type algorithm to deal with the likelihood's nastiness before it switches to the much faster standard BFGS algorithm (Broyden, 1970), performs reasonably well in all four applications discussed in this thesis.

1.1.3 Identification

Furthermore, some specific identification problems which may arise during the estimation of a finite mixture model should be briefly mentioned at this point. A more formal and extensive discussion of the identification issues in finite mixture models may be found in Frühwirth-Schnatter (2006). First, as equation (1.1) shows, any finite mixture model's density is obviously only identified up to an arbitrary permutation in the C different groups' labeling, i.e. the likelihood is the same when the groups' labels are interchanged. As long as the researcher is exclusively interested in the maximum likelihood estimates, this is of minor relevance. But as soon as she wants to use standard errors obtained by the bootstrap method (Efron and Tibshirani, 1993) or estimate the model from a Bayesian point of view by applying MCMC simulation techniques, the groups' labels will spontaneously change from one draw to the other and, consequently, so called label switching becomes a problem. Normally, label switching can be effectively prevented by imposing some straightforward identification restriction on the groups' labels such as $\pi_1 < \pi_2 < \ldots < \pi_C$, for example. Second, even after avoiding label switching, a finite mixture model with fully identified component densities and a data matrix, $\mathcal{X}$, which has full rank, may still be unidentified. As Grün and Leisch (2004) point out, in such a case lack of identification occurs when, intuitively speaking, the variables in $\mathcal{X}$ take on too few different realizations relative to the number of groups, C. Panel (a) in Figure 1.1

Figure 1.1: Identification of Finite Mixture Models



Following Grün and Leisch (2004).

illustrates such a situation for the simple example of a linear regression of an outcome, y, on a binary independent variable, x, with the two possible realizations x_1 and x_2. In this case, regardless of sample size, the data never contains sufficient information to identify a finite mixture specification with $C = 2$ components, i.e. the solution with the two straight regression lines is equally likely as the solution with the two dashed ones. But if x can take on a third value, x_3, the model is identified as shown in Panel (b). In panel data sets with several observations per individual, however, the model can also be identified by choosing individuals rather than observations as unit of classification. Thus, the often natural assumption that all observations of one individual must belong to the same group imposes additional structure on the model and ensures its identifiability. In Figure 1.1, this is depicted in the bottom Panel (c) where all data points from one individual are represented by the same symbol in the scatter plot. In practice, identification through the data's panel structure often turns out to be crucial for reliably fitting complex finite mixture regression models, even when they may be theoretically identified without such an additional structure.

Moreover, if the researcher wants to use a finite mixture model for classification purposes and has no theoretical *a priori* knowledge about the optimal number of behavioral types in the population, the question of how to choose C optimally turns out to be very difficult and remains, at least to some extent, still open (Frühwirth-Schnatter, 2006). It is well known that classical information criteria, such as the Akaike or Bayesian Information Criterion, tend to overestimate the number of groups and lead to a specification which overfits the data (Celeux and Soromenho, 1996). However, there are three different approaches which may provide an indication about the optimal number of classes: applying simulated likelihood ratio tests, approximating the model's marginal likelihood in a Bayesian framework (Houser et al., 2004), or minimizing a normalized entropy criterion (El-Gamal and Grether, 1995; Celeux and Soromenho, 1996). Besides relying on theoretical arguments about the number of behavioral types, the four applications presented in this thesis follow the latter approach and apply El-Gamal's Average Normalized Entropy, ANE, which is not only relatively inexpensive to compute but has also an intuitive interpretation.

1.2 Structure of This Thesis

This thesis comprises four independent applications of finite mixture regression models. The first three experimental studies are part of a comprehensive research project, funded by the Swiss National Science Foundation (SNSF), and discuss the identification and stability of two different behavioral types of decision makers in the domain of risk. The last essay applies a finite mixture model to the German Socio-Economic Panel (GSOEP) to segregate the share of altruists from the rest of the population which is assumed to be selfish.

Chapter 2 discusses the first contribution to the SNSF project. The past decades of experimental economic research have demonstrated that there is considerable heterogeneity in individual risk taking behavior, but little is known about the distribution of risk taking types. Together with my coauthors, Helga Fehr and Thomas Epper, I present a parsimonious characterization of risk taking behavior by estimating a finite mixture regression model for three different experimental data sets, two Swiss and one Chinese. In all three data sets participants' certainty equivalents are elicited for a large number of binary lotteries over real gains and losses. We find a robust segregation into two distinct behavioral types: In all three data sets, the choices of roughly 80 percent of the subjects exhibit significant deviations from rational probability weighting, consistent with a rank- and sign-dependent decision model such as proposed by prospect theory. 20 percent of the subjects, however, do not distort probabilities and behave essentially as expected value maximizers. Furthermore, the model cleanly segregates the individuals into these two groups and achieves a low entropy measure. Thus, to avoid aggregation bias in situations with strategic complementarity, researchers in applied economic modeling may consider using a mix of preference theories rather than assuming a representative decision maker.

In Chapter 3 we investigate how risk tolerance varies with stake size in the Chinese data set. On the one hand, this question is generally interesting for assesssing the external validity of experimental studies, which typically involve substantially lower stakes than do important risky decisions in real life. On the other hand, as the actual composition of the population may be decisive for market outcomes, it is important to know whether the segregation into two distinct behavioral types, as discussed in Chapter 2, remains stable when the lotteries' stakes are increased considerably. On the aggregate level, an increase in stakes from about an hourly wage to roughly a monthly income causes the individuals to behave relatively more risk averse for gains but not for losses, for which risk aversion remains essentially stable. When we control for latent individual heterogeneity by estimating the finite mixture regression model, the separation in one fifth expected utility theory types and four fifths prospect theory types turns out to be robust to increasing stakes. Moreover, the expected utility theory types still behave risk neutrally, whereas the other group's higher average degree of risk aversion is largely due to a shift in their probability weights. Since probability weights should, in theory, not be affected by stake size, this result for the majority group questions prospect theory as a proper descriptive model.

Chapter 4 therefore assesses the performance of prospect theory versus stochastic utility theory, a recent model for individual decision making under risk, which extends expected utility theory and assumes that individuals are prone to make random errors when computing a lottery's expected utility (Blavatskyy, 2007). For aggregate data the results are mixed: Notwithstanding its parsimony, stochastic expected utility theory clearly outperforms prospect theory in the Zurich data. In China, however, prospect theory seems to have a slight edge over the model based on stochastic expected utility theory. Nevertheless, if I take latent heterogeneity into account and estimate a finite mixture specification of the two theories, a consistent picture emerges. With a few exceptions, the individuals are segregated into the same two behavioral types as in Chapter 2. The risk neutral group's behavior is represented by stochastic expected utility theory, whereas the rest of the population, still, is best described by prospect theory. Hence, even if stochastic expected utility theory is an elegant and parsimonious approach, which generally describes aggregate choices well, it seems to fall short of prospect theory's descriptive power when individual heterogeneity is taken into account.

The final Chapter 5 extends the finite mixture approach to survey data. Even though several results from experimental economics suggest that the population is made up of different social preference types (Andreoni and Miller, 2002; Charness and Rabin, 2002), there exists barely any survey-based evidence on individuals exhibiting heterogeneous social preferences. In this study, Rainer Winkelmann and I investigate the prevalence and extent of altruism by examining the relationship between parents' and their adult children's subjective well-being in a sample extracted from the GSOEP. To segregate the share of parents with altruistic preferences from those who are assumed to be strictly selfish, we estimate a finite mixture extension of the ordered probit model. Furthermore, we apply a family-effects estimator to avoid biased estimators due to various potential sources of endogeneity. The estimate of the altruist's share among the population amounts to roughly 20 percent. Not only does this fraction of altruists coincide with a recent survey-based psychological study, but also the parents who get identified as altruists indeed make higher average transfer payments to their children.

Bibliography

ANDREONI, J. AND J. MILLER (2002): "Giving According to GARP: An Experimental Test of The Rationality of Altruism," *Econometrica*, 70, 737–753.

BLAVATSKYY, P. R. (2007): "Stochastic Expected Utility Theory," *Journal of Risk and Uncertainty*, 34, 259–286.

BROYDEN, C. (1970): "The Convergence of a Class of Double-Rank Minimization Algorithms," *Journal of the Institute of Mathematics and its Applications*, 6, 76–90.

CELEUX, G., D. CHAUVEAU, AND J. DIEBOLT (1996): "Stochastic Versions of the EM Algorithm: An Experimental Study in the Mixture Case," *Journal of Statistical Computation and Simulation*, 55, 287–314.

CELEUX, G. AND G. SOROMENHO (1996): "An Entropy Criterion for Assessing the Number of Clusters in a Mixture Model," *Journal of Classification*, 13, 195–212.

CHARNESS, G. AND M. RABIN (2002): "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117, 817–868.

CLARK, A., F. ETILE, F. POSTEL-VINAY, C. SENIK, AND K. V. DER STRAETEN (2005): "Heterogeneity in Reported Well-Being: Evidence from Twelve European Countries," *The Economic Journal*, 115, C118–C132.

DEMPSTER, A., N. LAIRD, AND D. RUBIN (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

EFRON, B. AND R. J. TIBSHIRANI (1993): *An Introduction to the Bootstrap*, Chapman and Hall/CRC.

EL-GAMAL, M. A. AND D. M. GRETHER (1995): "Are People Bayesian? Uncovering Behavioral Strategies," *Journal of the American Statistical Association*, 90, 1137–1145.

FEHR, E. AND J. R. TYRAN (2005): "Individual Irrationality and Aggregate Outcomes," *Journal of Economic Perspectives*, 19, 43–66.

Frühwirth-Schnatter, S. (2006): *Finite Mixture and Markov Switching Models*, Springer.

Grün, B. and F. Leisch (2004): "Bootstrapping Finite Mixture Models," in *Compstat 2004 — Proceedings in Computational Statistics*, Physica Verlag, Heidelberg, 1115–1122.

Haltiwanger, J. C. and M. Waldman (1985): "Rational Expectations and the Limits of Rationality: An Analysis of Heterogeneity," *American Economic Review*, 75, 326–340.

——— (1989): "Limited Rationality and Strategic Complements: The Implications for Macroeconomics," *Quarterly Journal of Economics*, 104, 463–483.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Heckman, J. and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271–3220.

Houser, D., M. Keane, and K. McCabe (2004): "Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence using a Bayesian Type Classification Algorithm," *Econometrica*, 72, 781–822.

McCutcheon, A. L. (1987): *Latent Class Analysis*, Sage Publications.

McLachlan, G. and D. Peel (2000): *Finite Mixture Models*, Wiley Series in Probabilities and Statistics.

Pearson, K. (1894): "Contributions to the Theory of Mathematical Evolution," *Philosophical Transactions of the Royal Society of London*, 185, 71–110.

Priebe, C. (1994): "Adaptive Mixtures," *Journal of the American Statistical Association*, 89, 796–806.

Render, R. A. and X. Walker (1984): “Mixture Densities, Maximum Likelihood and the EM Algorithm,” *SIAM Review*, 26, 195–239.

Stahl, D. O. and P. W. Wilson (1995): “On Players’ Models of other Players: Theory and Experimental Evidence,” *Games and Economic Behavior*, 10, 218–254.

Chapter 2

Risk and Rationality: Uncovering Heterogeneity in Probability Distortion

This chapter is joint work with Helga Fehr and Thomas Epper.

Another version of this chapter has been published as *SOI Working Paper 0705* and is currently under review at *Econometrica*.

2.1 Introduction

Risk is a ubiquitous feature of social and economic life. Many of our everyday choices, and often the most important ones, such as what trade to learn and where to live, involve risky consequences. While it has long been recognized that individuals differ in their risk taking attitudes, surprisingly little is known about the distribution of risk preferences in the population (for an exception see Dohmen et al. (2005)). Since preferences are one of the ultimate drivers of behavior, knowledge of the composition of risk attitudes is paramount to predicting economic behavior. Economic models often allow for heterogeneity, but this heterogeneity is usually confined to remain within the boundaries of the standard model of preferences, expected utility theory (EUT). The empirical evidence, however, reveals that heterogeneity in risk taking behavior is of a substantive kind, i.e. some people evaluate risky prospects consistently with EUT, whereas other people depart substantially from expected utility maximization (Hey and Orme, 1994). Moreover, it seems to be the case that rational decision makers revealing EUT-preferences constitute only a minority of the population. To improve descriptive performance a plethora of alternative theories have been developed. Unfortunately, no single best fitting model has been identified so far (Harless and Camerer, 1994; Starmer, 2000) and, depending on the individual, one or the other model fits better. This finding poses a serious problem for applied economics. What the modeler needs is a *parsimonious* representation of risk preferences that is empirically well grounded and robust, and not a host of different functionals. Providing such a parsimonious characterization of heterogeneity in risk taking behavior is the objective of this paper.

Our method is based on a literature on classifying individuals which has recently emerged in the social sciences. On the basis of statistical classification procedures, such as finite mixture regression models, investigators have tried to discover which decision rules people actually use when playing games or dealing with complex decision situations (El-Gamal and Grether, 1995; Houser et al., 2004; Houser and Winter, 2004). The finite mixture regression approach does not require fitting a model for each individual, which is - given the usual quality of choice data - frequently impossible. Instead, our approach reveals latent heterogeneity by estimating the fractions of distinct behavioral types and by endogenously assigning each individual to one specific type, characterized by a unique set of parameter values.

We apply such a finite mixture regression model to choice data from three different experiments, two of which were conducted in Zurich, Switzerland. The third experiment took place in Beijing, People's Republic of China. We analyze 452 subjects' decisions over real monetary gains and losses, which comprise a total of about 18,000 choices. All three experiments were designed in a similar manner and served to elicit certainty equivalents for binary lotteries. Using a flexible sign-dependent functional as basic behavioral model, we show the following results.

First, the estimation procedure renders a robust classification of risk taking behavior across all three data sets. Irrespective of time and place of the experiments, two distinct behavioral types of individuals are identified. Moreover, the proportions of these distinct types in their respective populations are practically equal in both the Swiss and the Chinese data sets and amount to roughly 20:80.

Second, almost all the experimental subjects are unambiguously assigned to one of the two distinct types. Measuring the quality of classification by the Average Normalized Entropy (El-Gamal and Grether, 1995), ambiguity of assignments amounts to less than 5% of maximum entropy, a value which is, to our knowledge, unequaled in the literature. Thus, we observe hardly any "ambiguous" types, i.e. individuals with a high probability (of say 0.4) of being one type *and* a high probability (of say 0.6) of being the other type are practically absent. This clean segregation suggests that the classification procedure is able to capture the distinctive characteristics of each behavioral type.

Third, without restricting parameter values *a priori*, we find that, in all three data sets, the minority types weight probabilities and evaluate value monetary outcomes nearly linearly. Consequently, this group of individuals can essentially be characterized as expected value maximizers. This result is particularly interesting in the light of Rabin's calibration theorem (Rabin, 2000), which shows that expected utility maximizers should be approximately risk neutral for small stakes typically encountered in laboratory experiments. Therefore, we label subjects belonging to this group of nearly risk neutral people as "EUT types".

Fourth, the majority of individuals, labeled as “CPT types”, are characterized by significant deviations from linear probability weighting, consistent with prospect theory. For all three prospect theory groups we obtain similar parameter values for the probability weighting functions and the value functions over losses, indicating considerable temporal and cross-cultural stability of preference parameters. For decisions over gains, however, Chinese behavior differs substantially from Swiss behavior. Overweighting of probabilities is more pronounced and the sensitivity to changes in probabilities is substantially lower for Chinese subjects, rendering the Chinese relatively more risk seeking for gains over a considerable range of probabilities. Thus, the finite mixture regression helps to better understand the nature of cross-cultural differences.

These results show that the classification procedure successfully uncovers latent heterogeneity in the population. If there is heterogeneity of a substantive kind, as the data suggest, basing predictions on a single preference theory is inappropriate and may lead to biased results. EUT preferences should be taken account of alongside prospect theory preferences, even if rational behavior constitutes only a minority in the population. As the literature on the role of bounded rationality under strategic complementarity and substitutability has shown (Haltiwanger and Waldman, 1985, 1989; Fehr and Tyran, 2005; Camerer and Fehr, 2006), the mix of rational and irrational actors may be decisive for aggregate outcomes. Depending on the nature of the strategic interdependence the behavior of even a minority of players may drive the aggregate outcome. Therefore, the mix of types in the population is a crucial variable in predicting market outcomes. Since the finite mixture regression provides a robust and reliable classification of individuals, the resulting estimates of group sizes and group-specific parameters may serve as valuable inputs for applied economics.

To the best of our knowledge, there is no previous study showing a nearly identical classification of risk preference types for three independent data sets. Related work by Harrison and colleagues (Harrison and Rutström, 2005; Harrison et al., 2005; Andersen et al., 2006) also applies finite mixture regressions to several experimental data sets, but decisively distinguishes itself from our analysis. Their estimation procedure sorts *choices*, irrespective by whom they were taken, by decision model, while we aim at classifying *individuals* by behavioral type. Since different types of individuals may behave the same way in certain decision situations, namely when probability weighting does not play much of a role, a classification of choices cannot reliably identify different individual types of risk taking behavior. Therefore, distinguishing EUT-consistent choices from CPT-consistent ones in these specific decision situations is nearly impossible. As a consequence, a classification of *choices* is bound to be highly ambiguous, even though there are clearly distinct types of decision makers. This may be one reason why Harrison and colleagues find a rather ambiguous classification of choices, while we unambiguously identify two distinct types of *decision makers*.

Moreover, a classification of choices depends on the mix of lotteries. If subjects face a large proportion of lotteries with gain probabilities in the vicinity of 0.4, the weight of EUT tends to be relatively high because of the ambiguity of classification for these lotteries. This dependence on the mix of lotteries renders a classification of choices rather arbitrary.

In addition, Harrison and colleagues *restrict* one decision model to be consistent with expected utility theory, whereas our estimation procedure assigns subjects to one of two *endogenously* defined types, one of which turns out to be essentially consistent with EUT-preferences. Thus, our results can be viewed as much stronger evidence in favor of EUT.

The paper is structured as follows. Section 2.2 describes the experimental design and procedures of the three experiments. The functional specification of the behavioral model and the finite mixture regression model are discussed in Section 2.3. Section 2.4 presents descriptive statistics of the data and the results of the classification procedure. Section 2.5 concludes.

2.2 Experimental Design

In the following section we describe the experimental setup and procedures. The experiments took place in Zurich in 2003 and 2006 as well as in Beijing in 2005. In Zurich, all subjects were recruited from the subject pool of the Institute for Empirical Research in Economics, which consists of students of all fields of the University of Zurich and the

Table 2.1: Differences in Experimental Design

	Zurich 03	Zurich 06	Beijing 05
Number of:			
Subjects	181	118	153
Lotteries	50	40	28
Observations	9,005	4,669	4,281
Procedure	computerized	computerized	paper and pencil
Framing	abstract and contextual	contextual	abstract and contextual

Swiss Federal Institute of Technology Zurich. In Beijing, subjects were recruited by flier distributed at the campus of Peking University and Tsinhua University. Since all three experiments are based on the same design principles, we will present the prototype experiment Zurich 2003 in detail (Fehr-Duda et al., 2006) and describe to what extent the other two experiments deviate from the prototype. The main distinguishing features of the different experiments are summarized in Table 2.1.

We elicited certainty equivalents for a large number of two-outcome lotteries. One half of the lotteries were framed as choices between risky and certain gains ("gain domain"), the other half were presented as choices between risky and certain losses ("loss domain"). For each decision in the loss domain, subjects were endowed with a specific monetary amount, which served to cover potential losses and equalized expected payoffs of corresponding gain and loss lotteries. In the Zurich 2003 and the Beijing experiments, 50% of the subjects were confronted with decisions framed in the standard gamble format. The other 50% of the subjects had to make choices framed in contextual terms, i.e. gains were represented as risky or sure investment gains, losses as repair costs and insurance premiums, respectively. The Zurich 2006 experiment was based on contextually framed lotteries only. In Zurich, outcomes x_1 and x_2 ranged from zero Swiss Francs to 150 Swiss Francs[1]. The payoffs in the Beijing 2005 experiment were commensurate with the compensation in Zurich and varied between 4 and 55 Chinese Yuan. Expected payoffs per subject amounted to approximately 31 Swiss Francs and 20 Chinese Yuan, respectively, which was considerably more than a local student assistant's hourly compensation, plus a show up fee of 10 Swiss Francs and

[1]At the time of the experiment one Swiss Franc equalled about 0.90 U.S. Dollars.

Table 2.2: Gain Lotteries Zurich 03 $(x_1, p; x_2)$

p	x_1	x_2	p	x_1	x_2	p	x_1	x_2
0.05	20	0	0.25	50	20	0.75	50	20
0.05	40	10	0.50	10	0	0.90	10	0
0.05	50	20	0.50	20	10	0.90	20	10
0.05	150	50	0.50	40	10	0.90	50	0
0.10	10	0	0.50	50	0	0.95	20	0
0.10	20	10	0.50	50	20	0.95	40	10
0.10	50	0	0.50	150	0	0.95	50	20
0.25	20	0	0.75	20	0			
0.25	40	10	0.75	40	10			

Outcomes x_1 and x_2 are denominated in Swiss Francs (CHF).

20 Chinese Yuan, thus generating salient incentives. Probabilities p of the lotteries' higher gain or loss x_1 varied from 5% to 95%. The gain lotteries for Zurich 2003 are presented in Table 2.2. The other two experiments essentially included a subset of these. The lotteries appeared in random order on a computer screen[2], in Beijing on paper.

In the computerized experiments, the screen displayed a decision sheet containing the specifics of the lottery under consideration and a list of 20 equally spaced certain outcomes, ranging from the lottery's maximum payoff to the lottery's minimum payoff, as shown in Figure 2.1[3]. The subjects had to indicate whether they preferred the lottery or the certain payoff for each row of the decision sheet. The lottery's certainty equivalent was calculated as the arithmetic mean of the smallest certain amount the subject preferred to the lottery and the following certain amount on the list, when the subject had, for the first time, reported preference for the lottery. For example, if the subject had decided as indicated by the small circles in Figure 2.1, her certainty equivalent would amount to 13.5 Swiss Francs.

Before subjects were permitted to start working on the experimental decisions, they had to correctly calculate the payoffs for two hypothetical choices. In the computerized experiments, there were two trial rounds to familiarize the subjects with the procedure. At the end of the experiment, one row number of one decision sheet was randomly selected for each subject, and the subject's choice in that row determined her payment. Subjects

[2]The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).
[3]The format of the decision sheet for the Beijing experiment was identical to the Zurich one.

Figure 2.1: Design of the Decision Sheet

Decision situation:
22

	Option A	**Your Choice:**				**Option B Guaranteed payoff amounting to:**
1		**A**		**o**	**B**	20
2		**A**		**o**	**B**	19
3		**A**		**o**	**B**	18
4		**A**		**o**	**B**	17
5		**A**		**o**	**B**	16
6		**A**		**o**	**B**	15
7	A profit of CHF 20 with	**A**		**o**	**B**	14
8		**A**	**o**		**B**	13
9	probability 75%	**A**	**o**		**B**	12
10		**A**	**o**		**B**	11
11	and a profit of CHF 0 with	**A**	**o**		**B**	10
12		**A**	**o**		**B**	9
13	probability 25%	**A**	**o**		**B**	8
14		**A**	**o**		**B**	7
15		**A**	**o**		**B**	6
16		**A**	**o**		**B**	5
17		**A**	**o**		**B**	4
18		**A**	**o**		**B**	3
19		**A**	**o**		**B**	2
20		**A**	**o**		**B**	1

OK

were paid in private afterward. The subjects could work at their own speed, the vast majority of them needed less than an hour to complete the experiment.

2.3 Econometric Model

This section discusses the specification of the finite mixture regression model, which allows controlling for latent heterogeneity in risk taking behavior in a parsimonious way. Estimating the finite mixture model yields the relative sizes of a pre-specified number of groups and the group-specific parameters of the underlying behavioral model. Moreover, as we use the Expectation Maximization algorithm (Dempster et al., 1977) to compute the maximum likelihood estimates of the model parameters, we obtain Bayesian updates for the probabilities of individual group membership. This procedure allows us to assign each individual to a specific group.

For the purpose of classifying subjects according to risk taking type, we need to specify three ingredients of the mixture model: the basic theory of decision under risk, the functional form of the decision model, and the specification of the error term.

The underlying theory of decision under risk should be able to accommodate a wide range of different behaviors. Sign- and rank-dependent models, such as cumulative prospect theory (CPT), capture two robust empirical phenomena: nonlinear probability weighting and loss aversion (Starmer, 2000). Therefore, a flexible approach, such as proposed by CPT, lends itself to describing risk taking behavior. Moreover, CPT nests EUT as special case. If there is a group of people, whose behavior can best be described by EUT, these individuals should be identified by the finite mixture regression as a unique group exhibiting the predicted behavior.

Suppose that there are C different types of individuals in the population. According to CPT, an individual belonging to a certain group $c \in \{1, \ldots, C\}$ values any binary gamble $\mathcal{G}_g = (x_{1g}, p_g; x_{2g})$, $g \in \{1, \ldots, G\}$, where $|x_{1g}| > |x_{2g}|$, by

$$v\left(\mathcal{G}_g\right) = v(x_{1g})w(p_g) + v(x_{2g})(1 - w(p_g)). \tag{2.1}$$

The function $v(x)$ describes how monetary outcomes x are valued, whereas the function $w(p)$ assigns a subjective weight to every outcome probability p. The gamble's certainty equivalent $\hat{ce}_g$ can then be written as

$$\hat{ce}_g = v^{-1}\left[v(x_{1g})w(p_g) + v(x_{2g})(1 - w(p_g))\right]. \tag{2.2}$$

In order to make CPT operational, we have to assume specific functional forms for the value function $v(x)$ and the probability weighting function $w(p)$. A natural candidate for $v(x)$ is a sign-dependent power functional

$$v(x) = \begin{cases} x^{\alpha} & \text{if } x \geq 0 \\ -(-x)^{\beta} & \text{otherwise,} \end{cases} \tag{2.3}$$

which can be conveniently interpreted and has turned out to be the best compromise between parsimony and goodness of fit in the context of prospect theory (Stott, 2006). For this specification of the value function, the existence of loss aversion can be inferred from the difference in the domain-specific curvatures. According to Tversky and Kahneman, loss aversion, in the sense that "losses loom larger than corresponding gains", is present if $v'(x) < v'(-x)$ for $x \geq 0$ (Tversky and Kahneman (1992), p.303). This is the case if the estimated β is significantly larger than α.

A variety of functional forms for modeling probability weights $w(p)$ have been proposed in the literature (Quiggin, 1982; Tversky and Kahneman, 1992; Prelec, 1998). We use the two-parameter specification suggested by Goldstein and Einhorn (1987) and Lattimore et al. (1992):

$$w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1-p)^{\gamma}}, \ \delta \geq 0, \ \gamma \geq 0. \tag{2.4}$$

We favor this specification because it has proven to account well for individual heterogeneity (Wu et al., 2004) and the parameters are nicely interpretable. The parameter γ largely governs the slope of the curve, whereas the parameter δ largely governs its elevation. The smaller the value of γ, the more strongly the probability weighting function departs from linear weighting. The larger the value of δ, the more elevated is the curve, ceteris paribus. Linear weighting is characterized by $\gamma = \delta = 1$. In a sign-dependent model, the parameters may take on different values for gains and for losses.

We now turn to the third step of model specification. In the course of the experiments, we measured risk taking behavior of individual $i \in \{1, \ldots, N\}$ by her certainty equivalents ce_{ig} for a set of different lotteries. Since CPT explains *deterministic* choice, we have to add an error term ϵ_{ig} in order to estimate the parameters of the model based on the elicited certainty equivalents. The observed certainty equivalent ce_{ig} can then be written as $ce_{ig} = \hat{ce}_g + \epsilon_{ig}$. There may be different sources of error, such as carelessness, hurry or inattentiveness, resulting in accidentally wrong answers (Hey and Orme, 1994). The Central Limit Theorem supports the assumption that the errors are normally distributed and simply add white noise.

Furthermore, we allow for three different sources of heteroskedasticity in the error variance. First, for each lottery the subjects have to consider 20 certain outcomes, which are equally spaced throughout the lottery's range $|x_{1g} - x_{2g}|$. Since the observed certainty equivalent ce_{ig} is calculated as the arithmetic mean of the smallest certain amount preferred to the lottery and the following certain amount, where the lottery is preferred, the error is proportional to the lottery range. Second, as the subjects may be heterogeneous with respect to their previous knowledge, their ability of finding the correct certainty equivalent as well as their attention span, we expect the error variance to differ by individual. Third, lotteries in the gain domain may be evaluated differently from the ones in the loss domain. Therefore, we allow for domain-specific variance in the error term. This yields the form $\sigma_{ig} = \xi_i |x_{1g} - x_{2g}|$ for the standard deviation of the error term distribution, where ξ_i denotes an individual domain-specific parameter. Note that the model allows to test for both individual-specific and domain-specific heteroskedasticity by either imposing the restriction $\xi_i = \xi$, or by forcing all the ξ_i to be equal in both decision domains. Both restrictions are rejected by their corresponding likelihood ratio tests in all three samples with p-values close to zero. Therefore, we control for all three types of heteroskedasticity in the estimation procedure.

Having discussed all the necessary ingredients, we now turn to the specification of the finite mixture regression model. The basic idea of the mixture model is assigning an individual's risk-taking choices to one of C different types of behavior, each characterized by a distinct vector of parameters $\theta_c = (\alpha_c, \beta_c, \gamma_c', \delta_c')'$[4]. We denote the proportions of these different types in the population by π_c. Given our assumptions on the distribution of the error term, the density of type c for the i-th individual can be expressed as

$$f\left(ce_i, \mathcal{G}; \theta_c, \xi_i\right) = \prod_{g=1}^{G} \frac{1}{\sigma_{ig}} \phi\left(\frac{ce_{ig} - \hat{ce}_g}{\sigma_{ig}}\right), \tag{2.5}$$

where $\phi(\cdot)$ denotes the density of the standard normal distribution. Since we do not know *a priori* to which group a certain individual belongs to, the proportions π_c are interpreted as probabilities of group membership. Therefore, each individual density of type c has to be weighted by its respective mixing proportion π_c, which, of course, is unknown and has to be estimated as well. Summing over all C components yields the individual's contribution to the model's likelihood L. The log likelihood of the finite mixture regression model is then given by

$$\ln L\left(\Psi; ce, \mathcal{G}\right) = \sum_{i=1}^{N} \ln \sum_{c=1}^{C} \pi_c f\left(ce_i, \mathcal{G}; \theta_c, \xi_i\right), \tag{2.6}$$

where the vector $\Psi = (\theta_1', \ldots, \theta_C', \pi_1, \ldots, \pi_{C-1}, \xi_1, \ldots, \xi_N)'$ summarizes all the parameters of the model which need to be estimated.

[4]The vectors γ_c and δ_c contain the domain-specific parameters for the slope and the elevation of the probability weighting functions.

The parameters are estimated by the iterative Expectation Maximization (EM) algorithm, which provides an additional feature: In each iteration, the algorithm calculates by Bayesian updating an individual's posterior probability τ_{ic} of belonging to group c. The posterior probabilities τ_{ic} represent a particularly valuable result of the estimation procedure. Not only do we obtain the probabilities of individual group membership, but we also have a method of judging the quality of the classification at our disposal. If all the τ_{ic} are either close to zero or one, all the individuals are unambiguously assigned to one specific group. The τ_{ic} can be used to calculate a summary measure of ambiguity, such as the Average Normalized Entropy (El-Gamal and Grether, 1995), in order to gage the extent of dubious assignments. If classification has been successful, i.e. if genuinely distinct types have been identified, we should observe a low measure of entropy.

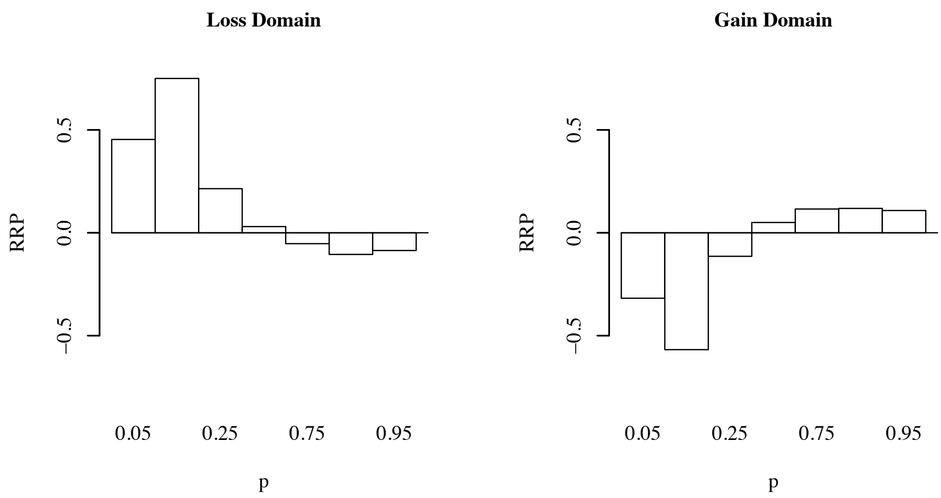
The resulting τ_{ic} also provide a basis for discriminating between models with differing numbers of types. Since the finite mixture regression model is defined over a pre-specified number of groups, the researcher needs a criterion for assessing the correct number of groups. In the context of mixture models, classical criteria, such as the Akaike Information Criterion AIC or the Bayesian Information Criterion BIC, are not suited for this purpose (Celeux and Soromenho, 1996). Celeux proposes to use entropy criteria based on the posterior probabilities of group assignment τ_{ic} instead, which are shown to perform better than the classical criteria.

Various problems may be encountered when maximizing the likelihood function of a finite mixture regression model and, therefore, a customized estimation procedure has to be used, which can adequately deal with these problems. Details of the estimation procedure, written in the R environment (R Development Core Team, 2006), are discussed in the Appendix.

2.4 Results

In the following section we present the results of the finite mixture regressions after describing observed risk taking behavior. First, we discuss the distributions of distinct risk taking types emerging in each of the three data sets and document the cleanness

Figure 2.2: Median Relative Risk Premia Zurich 2003



and robustness of individuals' segregation to types. Furthermore, each distinct type of individuals is characterized by the estimated behavioral parameter values, whereby we also address the issue of cross-cultural differences. Finally, we comment on the stability of classification with respect to model specification.

At the level of observed data, risk taking behavior can be conveniently summarized by relative risk premia $RRP = (ev - ce)/|ev|$, where ev denotes the expected value of a lottery's payoff and ce stands for its certainty equivalent. $RRP > 0$ indicates risk aversion, $RRP < 0$ risk seeking, and $RRP = 0$ risk neutrality. In the context of EUT, risk preferences are captured solely by the curvature of the utility function, which in turn determines the sign of relative risk premia. Therefore, the sign of RRP should be independent of p, the probability of the more extreme lottery outcome. In Figures 2.2 through 2.4, median risk premia sorted by p show a systematic relationship between RRP and p, however: In all three data sets subjects' choices display a fourfold pattern, i.e. they are risk averse for low-probability losses and high-probability gains, and they are risk seeking for low-probability gains and high-probability losses. Therefore, at a first glance, average behavior is adequately described by a model such as CPT rather than EUT. The median RRPs gloss over an important feature of the data, however: As the following discussion shows, there is substantial latent heterogeneity in risk taking behavior, which

Figure 2.3: Median Relative Risk Premia Zurich 2006

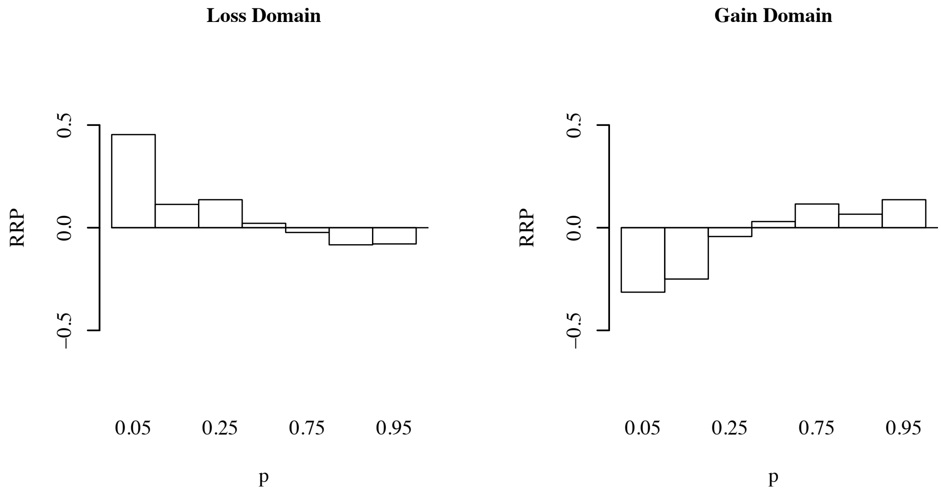


Figure 2.4: Median Relative Risk Premia Beijing 2005

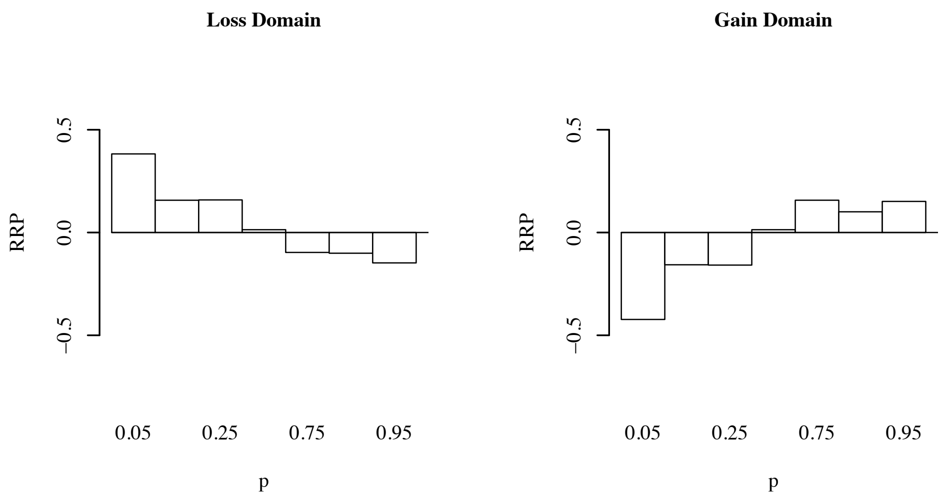


Table 2.3: Average Normalized Entropy, $C = 2$

	Zurich 03	Zurich 06	Beijing 05	Pooled
ANE	0.049	0.033	0.031	0.035

is uncovered by the finite mixture regressions.

2.4.1 Clean and Robust Segregation of Behavioral Types

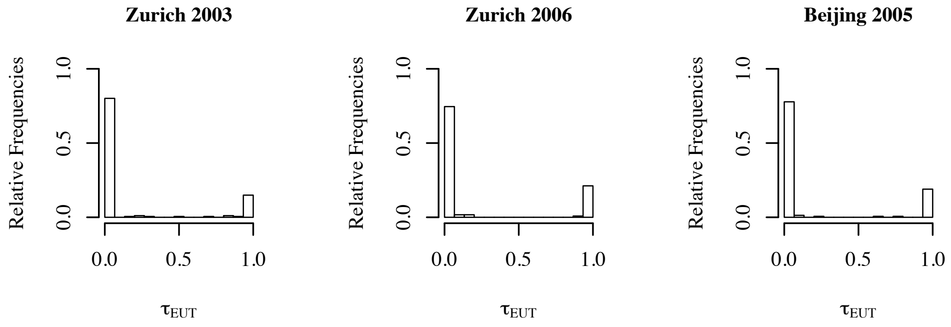
In order to be of value to applied economics, a classification of risk taking behavior should meet two conditions: First, it should be clean, i.e. all the individuals should be clearly associated with one specific risk taking type. Second, the classification should be robust across different experiments. Regarding the first condition, entropy criteria, based on the posterior probabilities of group assignment, can be used to evaluate the quality of classification. One such measure is the Average Normalized Entropy ANE (El-Gamal and Grether, 1995), defined as

$$ANE = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\tau_{ic}\log_C\left(\tau_{ic}\right), \tag{2.7}$$

for C groups and N individuals. This measure has the nice feature of lying within $[0, 1]$, as it is normalized by taking $\log_C$. If all the τ_{ic} are equal to zero or one, $ANE = 0$. In this case, all the individuals can be perfectly assigned to one of the different behavioral groups. $ANE = 1$ reflects maximum entropy, i.e. all the τ_{ic} are equal to $1/C$. Such a result would indicate that group membership is totally ambiguous and that categorization has failed. If this were the case, the model's assumption that there is a specific number of distinct types in the population could be refuted and, thus, using a finite mixture regression model would be inappropriate.

Since the finite mixture regression model provides a classification of individuals with respect to a pre-specified number of types, this number has to be assumed *a priori*. A natural starting point is to assume $C = 2$ and search for two distinct types, as previous empirical evidence suggests that there is a mix of EUT types and Non-EUT types in the population. The ANE for the two-group classifications estimated from our data are

Figure 2.5: Distribution of Posterior Probability of Assignment to EUT, τ_{EUT}



displayed in Table 2.3. All three classifications exhibit an average entropy of less than 5% of the maximal entropy of one, which is an extremely low degree of ambiguity by any standard. In their experiment on Bayesian learning, for instance, El-Gamal and Grether (1995) find the average entropy to lie between 11% and 38%, which they interpret to be "quite small". When pooling all three data sets, ANE amounts to 3.5%, underscoring the clean segregation of individuals into two distinct risk taking types even in a culturally diverse subject pool such as ours. These low values of ANE in our analysis indicate that nearly all the individuals can be unambiguously assigned to one distinct type of risk taking behavior.

The high quality of the two-group classifications can also be inferred directly from the distributions of the individuals' posterior probabilities of group assignment. In Figure 2.5, τ_{EUT} denotes the posterior probability of belonging to the first group, which can indeed be characterized, as we will demonstrate below, as expected utility maximizers. As the distributions of τ_{EUT} show, the individuals' posterior probabilities of behaving consistently with EUT are either close to one or close to zero for practically all the individuals in all three data sets, indicating an extremely clean segregation of subjects to types. Our result is quite remarkable as it substantiates that there are two distinct types in the population, be it Swiss, Chinese or culturally mixed, as in the pooled data set, and not a continuity of heterogeneous types. And it also shows that the underlying behavioral model provides a sound basis of discriminating between them.

With respect to the second criterion, robustness of classification, Figure 2.5 illustrates the probably most striking result of our study, namely similar distributions of types across all three data sets. In all three histograms of Figure 2.5, there are about four times as many individuals with τ_{EUT} close to zero, compared to individuals with τ_{EUT} close to one. This finding is mirrored by the estimates of the mixing proportions π_c. Table 2.4 displays, for the individual and the pooled data sets, the group-specific parameter estimates of the finite mixture regression model and their standard errors, obtained by the bootstrap method with $4,000$ replications (Efron and Tibshirani, 1993). In all the cases, estimates of the first groups' fractions amount to about 20% and, consequently, to about 80% for the second group. Moreover, the 95%-confidence intervals for the estimates of π_c for all three data sets overlap. Therefore, the classification is not only unambiguous, but also results in roughly equal proportions of both types across our data sets, demonstrating that classification is robust to experimental design, time and place.

This finding leads us to the next question. Do the respective types identified in each data set also exhibit similar patterns of behavior? This question will be addressed in the following two sections, dedicated to the characterization of the two endogenously defined types of behavior.

2.4.2 Characterization of the Minority Type

The first type of individuals encompasses about 20% of the subjects in all three data sets, thus constituting the minority types. Risk taking behavior is represented by the parameter estimates of the value functions and probability weighting functions. Concerning the latter model component, Table 2.4 displays almost identical parameter estimates across all three data sets. Without having imposed any restrictions on the parameters, we find that the minority groups' probability weighting functions are roughly linear, as the parameter estimates for both γ and δ are close to one. Since the probability weights are a nonlinear combination of these parameters, inference needs to be based on γ and δ jointly. It could well be the case that one of the two parameter estimates is significantly different from one, but the confidence band of the curve still includes the diagonal. Therefore, we constructed the 95%-confidence bands for the probability weighting curves by the percentile bootstrap method. Figures 2.6, 2.7, and 2.8 contain the graphs of the type-

Table 2.4: Classification of Behavior

	EUT Types				CPT Types			
Parameters	ZH 03	ZH 06	BJ 05	Pooled	ZH 03	ZH 06	BJ 05	Pooled
π	0.176	0.224	0.201	0.195	0.824	0.776	0.799	0.805
	(0.022)	(0.026)	(0.020)	(0.012)	(0.022)	(0.026)	(0.020)	(0.012)
Gains								
α	0.983	0.989	1.083	0.984	1.056	0.901	0.379	0.940
	(0.012)	(0.018)	(0.103)	(0.011)	(0.021)	(0.026)	(0.105)	(0.013)
γ	0.952	0.945	0.911	0.943	0.414	0.425	0.242	0.375
	(0.014)	(0.020)	(0.034)	(0.019)	(0.015)	(0.015)	(0.014)	(0.009)
δ	0.907	0.909	0.889	0.910	0.846	0.862	1.335	0.930
	(0.012)	(0.019)	(0.054)	(0.011)	(0.021)	(0.028)	(0.074)	(0.012)
Losses								
β	1.009	1.014	1.020	1.015	1.108	1.121	1.156	1.140
	(0.017)	(0.024)	(0.087)	(0.012)	(0.027)	(0.047)	(0.108)	(0.018)
γ	0.871	0.953	0.948	0.948	0.417	0.452	0.306	0.397
	(0.042)	(0.020)	(0.040)	(0.023)	(0.016)	(0.014)	(0.013)	(0.010)
δ	0.966	1.049	1.066	1.070	1.021	1.060	0.925	0.988
	(0.059)	(0.033)	(0.066)	(0.024)	(0.027)	(0.044)	(0.054)	(0.015)
$\ln L$	20,493	11,336	10,244	41,811				
Parameters	375	249	319	917				
Observations	9,005	4,669	4,281	17,955				

Standard errors (in parentheses) are based on the percentile bootstrap method with 4,000 replications.
Parameters include additional estimates for $\hat{\xi}_i$ for domain- and individual-specific error variances.
ZH stands for Zurich, BJ for Beijing.

specific probability weighting functions for each decision domain. The gray dotted lines correspond to the estimated curves for the first type, referred to as “EUT type”, and the gray dashed lines delimit their respective confidence bands. For both gains and losses, the confidence bands for the first type in fact include the diagonal over a wide range of probabilities, demonstrating high conformity with linear probability weighting. Where the confidence bands do not include the diagonal, the curves still lie extremely close to linear weighting. In sum, in all three data sets, we find the first behavioral type to exhibit near linear probability weighting.

Concerning the second model component, the valuation of monetary outcomes, the estimated parameters α and β also display a high degree of conformity. As can be inferred from the bootstrapped standard errors in Table 2.4, the 95%-confidence intervals of each

Figure 2.6: Probability Weighting Functions Zurich 2003

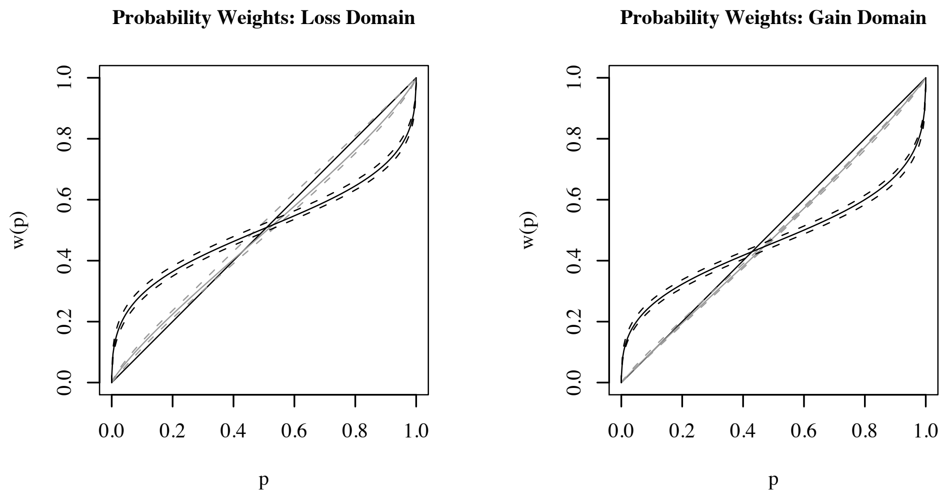


Figure 2.7: Probability Weighting Functions Zurich 2006

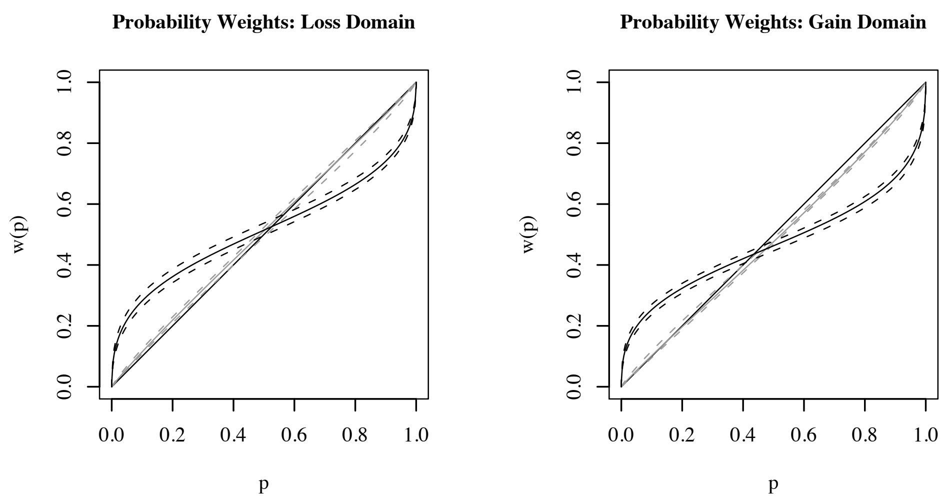
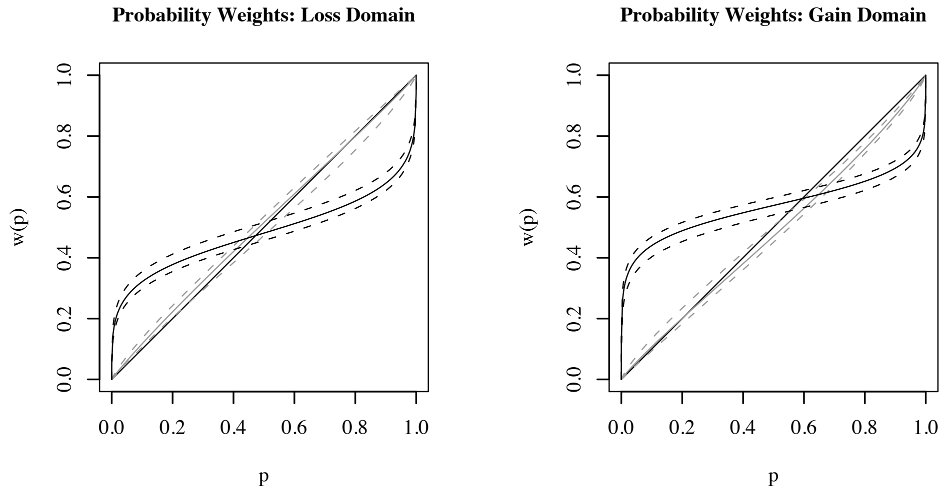


Figure 2.8: Probability Weighting Functions Beijing 2005



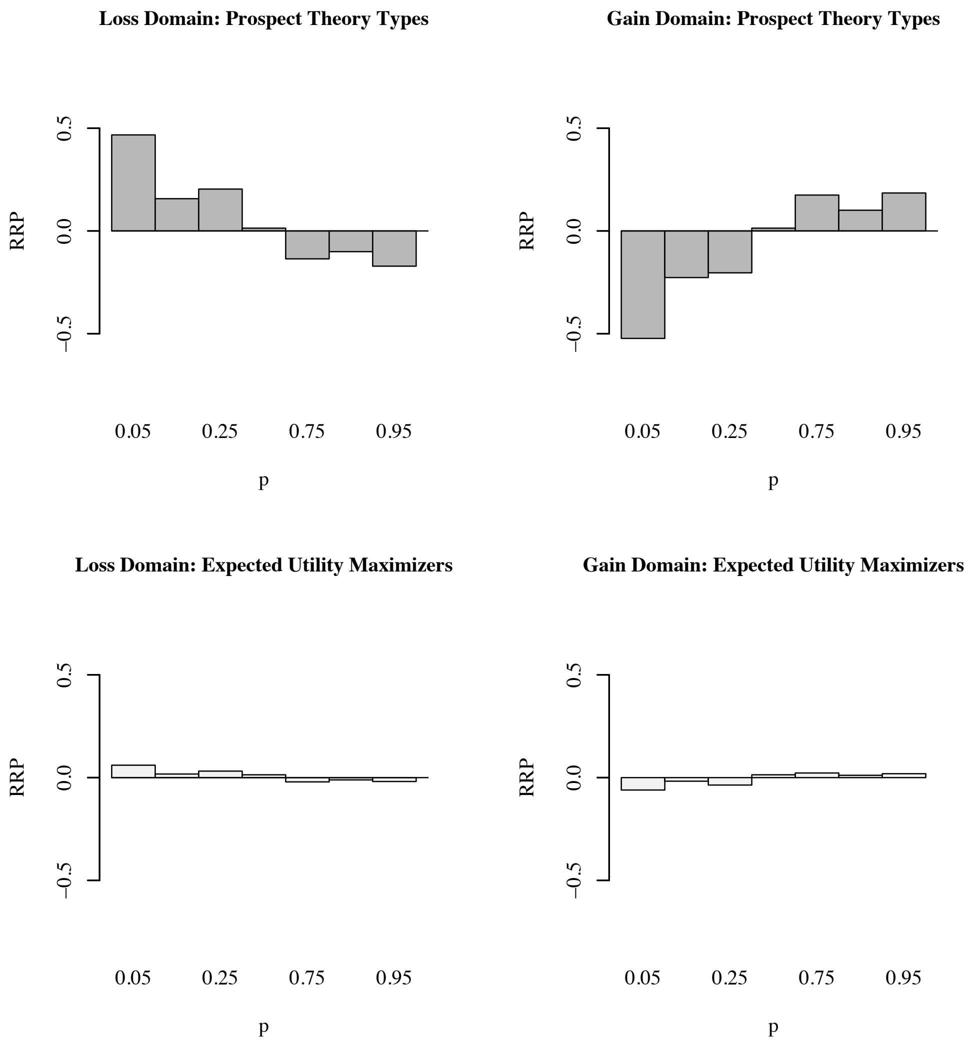
single curvature estimate contains the value of one, implying that the hypothesis of linear value functions cannot be rejected. Together with near linear probability weighting, this result justifies regarding the first type of individuals as largely consistent with expected value maximization, and therefore EUT.

The discriminatory power of the classification procedure can also be traced at the behavioral level. After assigning the subjects to one of the two types based on their τ_{ic}, the observed relative risk premia can be broken down by type as depicted in Figure 2.9, exemplary for the Chinese data set. As can be seen, median RRP of the Chinese EUT types are close to zero, reflecting near risk neutral behavior in accordance with expected value maximization. A similar picture can be shown to emerge for the Zurich 2003 and Zurich 2006 data sets. These findings document that not only the mix of behavioral types, but also the characterization of the minority type as essentially consistent with EUT is robust to experimental design, time and place.

2.4.3 Characterization of the Majority Type

The following results characterize the second, much larger, groups of individuals. As already suggested by the observed average patterns of risk attitudes, depicted in Figures

Figure 2.9: Median Relative Risk Premia by Type Beijing 2005



2.2 through 2.4, majority behavior exhibits a substantial dependence on the level of probability. The majority types' probability weighting curves are pictured as black lines in Figures 2.6, 2.7, and 2.8. The solid lines correspond to the estimated curves, the dashed lines mark the corresponding 95%-confidence bands. For both gains and losses, all three figures show inverted S-shaped probability weighting functions. Consequently, we label these individuals as "CPT types". However, CPT individuals do not display as uniform a behavior as the EUT individuals: Depending on decision domain, the parameter estimates in Table 2.4 reveal substantial cultural differences.

In the gain domain, the Swiss probability weighting curves exhibit the familiar shape, i.e. intersection with the diagonal at probabilities of about 0.4, whereas the Chinese probability weighting function is much more elevated than the Swiss ones, implying substantially more optimistic weighting of gain probabilities (estimated $\delta = 1.335$ versus 0.846 and 0.862, respectively). The Chinese probability weighting function is also considerably flatter in the middle part than the Swiss curves, which indicates a lower sensitivity towards changes in probabilities (estimated $\gamma = 0.242$ versus 0.414 and 0.425, respectively). Similarly, cultural differences are evident in the estimated value function parameters. Contrary to the slightly concave or convex Swiss curvatures, the Chinese value function is distinctly concave (estimated $\alpha = 0.38$).

In the loss domain, however, cultural differences are almost absent. The value function parameters β as well as the elevation parameters δ are estimated to be of the same orders of magnitude in Switzerland and in China. The Chinese probability weighting curve for losses departs more strongly from linearity, however, as the estimated slope parameter is significantly smaller than the corresponding Swiss values (estimated $\gamma = 0.306$ versus 0.417 and 0.452, respectively). This insensitivity to changes in probabilities seems to be a general feature of Chinese behavior.

The difference detected in the value function curvatures also has a bearing on loss aversion. As discussed in Section 2.3, the existence of loss aversion can be inferred from the difference in the domain-specific curvatures. If the estimated β is larger than α, "losses loom larger than corresponding gains". This is clearly the case for Zurich 06, but not for Zurich 03. Substantial loss aversion is also present in the Beijing 05 data set, another feature of cross cultural differences.

Tracing behavior of the CPT types at the level of observed RRP in Figure 2.9, we find a pronounced fourfold pattern of Chinese risk attitudes, with more extreme departures from risk neutrality than the aggregate risk premia in Figure 2.4. As before, a similar picture can be shown to emerge for the Zurich 2003 and Zurich 2006 data. This finding demonstrates that aggregate data underestimate the true extent of the CPT types' probability distortions.

2.4.4 Robustness to Model Specification

The final part of our analysis concerns questions of model specification. First, we discuss the issue of the correct number of groups in the finite mixture regression model. Second, we address an essential issue for prospect theory, namely, the assumption on the reference point implicit in our approach.

So far our results are based on the assumption of two distinct types, derived from the hypothesized co-existence of EUT and Non-EUT decision makers. In some applied contexts, it may be interesting to see what happens to classification when three groups are allowed for. If three types rendered a better characterization of risk taking behavior than two types, the Average Normalized Entropy should be smaller for $C = 3$ than for $C = 2$. Given the extremely low degree of entropy in our two-group classifications, an improvement in entropy, when three groups are assumed, seems hardly possible, however. Table 2.5 shows that ANE is indeed smaller for the two-group classifications than for the three-group classifications in all the cases. Moreover, we ascertained that relative group sizes as well as group membership of the EUT groups remain stable, when three-group models are estimated. Thus, regardless of whether $C = 2$ or $C = 3$ is assumed, the percentage of EUT types in the overall population amounts to approximately 20%. The CPT groups, however, get subdivided into two different CPT types, each characterized by a specific variety of nonlinear probability weighting, as there is some heterogeneity within the original CPT groups. Evidently, as the higher entropy measures for $C = 3$ in Table 2.5 show, these newly emerging types do not differ from each other as distinctly as do the overall CPT groups from EUT, since a number of CPT individuals cannot be clearly associated with one type, but are more suitably characterized by a mix of both types of CPT.

The second issue of interest concerns the question of reference point. Since prospect theory is silent on the subject of what “gains” and “losses” actually stand for, subjects’ reference points might be different from the ones the experimenter tries to induce, i.e. the endowment. For instance, people may not evaluate gambles in isolation, but integrate the prospective outcomes with their wealth or consumption spending. Therefore, we re-estimated the model with the value function being defined over the sum of the prospective

Table 2.5: Average Normalized Entropy by Number of Groups C

Groups	Zurich 03	Zurich 06	Beijing 05	Pooled
$C = 2$	0.049	0.033	0.031	0.035
$C = 3$	0.052	0.034	0.049	0.072

lottery outcome and an additional type-specific background parameter k, such that $v(x) = (x+k)^{\alpha}$ over gains and *mutatis mutandis* over losses. When such an endogenous reference point is included in the model, all our main results remain unchanged: In the extended model, the distributions of distinct behavioral types amount to roughly 20:80 in all three data sets. Segregation is extremely clean just as in the original model, and both EUT- and CPT-group memberships remain unchanged, with only 2 (!) of the 452 individuals not assigned to their original type. These findings confirm that there are two types of decision makers, consistent with EUT and CPT, respectively, and that classification is robust to model specification.

2.5 Concluding Remarks

We conducted three experiments based on the same design principles and applied a finite mixture regression model to the resulting data. For all three data sets a coherent picture emerges. Irrespective of composition of tasks, framing of the decisions, time and place of the experiments, we find an equal mix of two distinct groups. The classification procedure performs extremely well, resulting in less than 5% of the maximal Average Normalized Entropy, which means that almost all the individuals are reliably assigned to either one of the two distinct types. As it turns out, it is predominantly subjects' proneness to distorting stated probabilities that defines their behavioral type.

The first group comprises about 20% of the subjects, be they Swiss or Chinese, whose behavior can essentially be characterized by near linear probability weighting. Moreover, value function curvature estimates are not statistically different from linearity, implying near risk neutral behavior in line with the prediction of Rabin's calibration theorem. Consequently, we label this group of individuals EUT types.

The second group, encompassing approximately 80% of the subjects, can be classified as prospect theory types exhibiting an inverted S-shaped probability weighting function. In contrast to the EUT types, the overall group of CPT individuals, while sharing a general proneness to nonlinear probability weighting, is more heterogeneous. We find cultural differences particularly in choices over gains. While in the loss domain the behavior of the CPT types can be described by remarkably similar parameter values, Chinese CPT subjects tend to weight gain probabilities much more optimistically than do Swiss CPT subjects. The Chinese are also less responsive to changes in probabilities and display a substantial degree of loss aversion. When we estimate risk premia over a comparable range of outcomes, we predict the Chinese to be more risk seeking than the Swiss for gains of low and medium probability. Previous studies show that Chinese respondents indeed are relatively more risk seeking on average than Western ones (Kachelmeier and Shehata, 1992; Hsee and Weber, 1999). These results are consistent with our estimates and can be explained predominantly by the specific shape of the Chinese probability weighting function.

When we started this project we were quite confident that we would find a considerable fraction of expected utility maximizers. What really surprised us is the robust percentage of EUT types, even across two so different cultures as the Swiss and Chinese. This consistent magnitude of the EUT groups lends support to prior evidence by Hey and Orme (1994) and Lattimore et al. (1992). These near rational actors constitute a non-negligible proportion of the population whose behavior, depending on the nature of the strategic environment, may be decisive for aggregate outcomes. The existence of a robust share of near rational actors suggests to use a mix of preference theories for modeling behavior rather than a single theory, which would yield systematically biased results. Moreover, for the majority of subjects, prospect theory adequately describes behavior, but the parameter estimates exhibit culture-specific values. Researchers should take this evidence into account when constructing, estimating, and applying models of choice under risk.

2.6 Appendix: Estimation of the Finite Mixture Regression Model

As it is generally the case in finite mixture models, direct maximization of the log likelihood function

$$\ln L\left(\Psi; ce, \mathcal{G}\right) = \sum_{i=1}^{N} \ln \sum_{c=1}^{C} \pi_c \, f\left(ce_i, \mathcal{G}; \theta_c, \xi_i\right) \tag{2.8}$$

may encounter several problems, even if it is in principle feasible (for a general treatise see for example McLachlan and Peel (2000)). First, the highly non-linear form of the log likelihood causes the optimization algorithm to be rather slow or even incapable of finding the maximum. Second, the likelihood of a finite mixture model is often multimodal and therefore we have no guaranty that a standard optimization routine will converge towards the global maximum rather than to one of the local maxima.

However, if individual group membership were observable and indicated by $t_{ic} \in \{0, 1\}$ the individual contribution to the likelihood function would be given by

$$\tilde{\ell}\left(\Psi_i; ce_i, \mathcal{G}, t_i\right) = \prod_{c=1}^{C} \left[\pi_c \, f\left(ce_i, \mathcal{G}; \theta_c, \xi_i\right)\right]^{t_{ic}} \tag{2.9}$$

By using the above formulation and taking logarithms, the complete-data log likelihood function

$$\ln \tilde{L}\left(\Psi; ce, \mathcal{G}, t\right) = \sum_{i=1}^{N} \sum_{c=1}^{C} t_{ic} \left[\ln \pi_c + \ln f\left(ce_i, \mathcal{G}; \theta_c, \xi_i\right)\right] \tag{2.10}$$

would follow directly. As relative group sizes sum up to one, their maximum likelihood estimates, $\hat{\pi}_c = 1/N \sum_{i=1}^{N} t_{ic}$, would be given analytically by the relative number of individuals in the respective group. Furthermore, the maximum likelihood estimates of the group-specific parameters could be obtained separately in each group by numerically maximizing the corresponding joint density function which would simplify the optimization problem considerably.

The EM algorithm proceeds iteratively in two steps, E and M, while it treats the unobservable t_{ic} as missing data. In the E-step of the $(k+1)$-th iteration the expectation of the complete-data log likelihood $\tilde{L}$, given the actual fit of the data $\Psi^{(k)}$, is computed. This yields, according to Bayes' law, the posterior probabilities of individual group membership

$$\tau_{ic}\left(ce_i, \mathcal{G}; \Psi_i^{(k)}\right) = \frac{\pi_c^{(k)} f\left(ce_i, \mathcal{G}; \theta_c^{(k)}, \xi_i^{(k)}\right)}{\sum_{m=1}^{C} \pi_m^{(k)} f\left(ce_i, \mathcal{G}; \theta_m^{(k)}, \xi_i^{(k)}\right)} \tag{2.11}$$

which replace the unknown indicators of individual group membership, t_{ic}. Given these posterior probabilities $\tau_{ic}\left(ce_i, \mathcal{G}; \Psi_i^{(k)}\right)$, the complete-data log likelihood, $\tilde{L}$, is maximized in the following M-step which yields the updates of the model parameters,

$$\pi_c^{(k+1)} = \frac{1}{N} \sum_{i=1}^{N} \tau_{ic}\left(ce_i, \mathcal{G}; \Psi_i^{(k)}\right), \tag{2.12}$$

and

$$\begin{aligned} &\left(\theta_1^{(k+1)}, \ldots, \theta_C^{(k+1)}, \xi_1^{(k+1)}, \ldots, \xi_N^{(k+1)}\right) = \\ &\qquad \underset{\theta_1,\ldots,\theta_C,\xi_1,\ldots,\xi_N}{\arg\max} \sum_{i=1}^{N} \sum_{m=1}^{C} \tau_{im}\left(ce_i, \mathcal{G}; \Psi_i^{(k)}\right) \ln f\left(ce_i, \mathcal{G}; \theta_m^{(k)}, \xi_i^{(k)}\right). \end{aligned} \tag{2.13}$$

As Dempster et al. (1977) show, the likelihood never decreases from one iteration to the next, i.e. $L\left(\Psi^{(k+1)}; ce, \mathcal{G}\right) \geq L\left(\Psi^{(k)}; ce, \mathcal{G}\right)$, which makes the EM algorithm converge monotonically towards the nearest maximum of the likelihood function regardless whether this maximum is global or just local. In the Zurich 2003 data set, we therefore needed to apply a stochastic extension, the Simulated Annealing Expectation Maximization (SAEM) algorithm proposed by Celeux et al. (1996), in order to overcome the EM algorithm's tendency to converge towards local maxima. In each iteration, there is a non-zero probability that the SAEM algorithm leaves the current optimization path and starts over in a different region of the likelihood function which results in much higher chances of finding the global maximum. But this robustness against multimodality of the objective function comes at the cost of much higher computational demands.

As the EM algorithm is computationally highly demanding, even in its basic form, and tends to become tediously slow when close to convergence our estimation routine relies on a hybrid estimation algorithm (Render and Walker, 1984): It first uses either the EM or the SAEM algorithm and takes advantage of their robustness before it switches to the direct maximization of the log likelihood by the much faster BFGS algorithm.

The estimation routine in this form turned out to be efficient and robust as it reliably converged towards the same maximum likelihood estimates regardless of the randomly chosen start values.

Bibliography

Andersen, S., G. W. Harrison, and E. E. Rutström (2006): "Choice Behavior, Asset Integration and Natural Reference Points," *Department of Economics, College of Business Administration, University of Central Florida, Working Paper*, 06-07.

Camerer, C. F. and E. Fehr (2006): "When Does Economic Man Dominate Social Behavior?" *Science*, 311, 47–52.

Celeux, G., D. Chauveau, and J. Diebolt (1996): "Stochastic Versions of the EM Algorithm: An Experimental Study in the Mixture Case," *Journal of Statistical Computation and Simulation*, 55, 287–314.

Celeux, G. and G. Soromenho (1996): "An Entropy Criterion for Assessing the Number of Clusters in a Mixture Model," *Journal of Classification*, 13, 195–212.

Dempster, A., N. Laird, and D. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2005): "Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey," *Institute for the Study of Labor (IZA), Bonn (Germany), Discussion Paper*, 1730.

Efron, B. and R. J. Tibshirani (1993): *An Introduction to the Bootstrap*, Chapman and Hall/CRC.

El-Gamal, M. A. and D. M. Grether (1995): "Are People Bayesian? Uncovering Behavioral Strategies," *Journal of the American Statistical Association*, 90, 1137–1145.

Fehr, E. and J. R. Tyran (2005): "Individual Irrationality and Aggregate Outcomes," *Journal of Economic Perspectives*, 19, 43–66.

Fehr-Duda, H., M. de Gennaro, and R. Schubert (2006): "Gender, Financial Risk, and Probability Weights," *Theory and Decision*, 60, 283–313.

Fischbacher, U. (2007): "z-Tree: Zurich Toolbox for Readymade Economic Experiments," *Experimental Economics*, 171–178.

Goldstein, W. and H. Einhorn (1987): "Expression Theory and the Preference Reversal Phenomena," *Psychological Review*, 94, 236–254.

Haltiwanger, J. C. and M. Waldman (1985): "Rational Expectations and the Limits of Rationality: An Analysis of Heterogeneity," *American Economic Review*, 75, 326–340.

——— (1989): "Limited Rationality and Strategic Complements: The Implications for Macroeconomics," *Quarterly Journal of Economics*, 104, 463–483.

Harless, D. W. and C. F. Camerer (1994): "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica*, 62, 1251–1290.

Harrison, G. W., S. J. Humphrey, and A. Verschoor (2005): "Choice Under Uncertainty in Developing Countries," *Department of Economics, College of Business Administration, University of Central Florida, Working Paper*, 0529.

Harrison, G. W. and E. E. Rutström (2005): "Expected Utility Theory and Prospect Theory: One Wedding and A Decent Funeral," *Department of Economics, College of Business Administration, University of Central Florida, Working Paper*, 0518.

Hey, J. D. and C. Orme (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62, 1291–1326.

Houser, D., M. Keane, and K. McCabe (2004): "Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence using a Bayesian Type Classification Algorithm," *Econometrica*, 72, 781–822.

Houser, D. and J. Winter (2004): "How do behavioral assumptions affect structural inference?" *Journal of Business and Economic Statistics*, 22, 64–79.

Hsee, C. K. and E. U. Weber (1999): “Cross-National Differences in Risk Preferences and Lay Predictions,” *Journal of Behavioral Decision Making*, 12, 165–179.

Kachelmeier, S. J. and M. Shehata (1992): “Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence from the People’s Republic of China,” *American Economic Review*, 82, 1120–1141.

Lattimore, P. K., J. R. Baker, and A. D. Witte (1992): “The Influence of Probability on Risky Choice,” *Journal of Economic Behavior and Organization*, 17, 377–400.

McLachlan, G. and D. Peel (2000): *Finite Mixture Models*, Wiley Series in Probabilities and Statistics.

Prelec, D. (1998): “The Probability Weighting Function,” *Econometrica*, 3, 497–527.

Quiggin, J. (1982): “A Theory of Anticipated Utility,” *Journal of Economic Behavior and Organization*, 3, 323–343.

R Development Core Team (2006): “R: A Language and Environment for Statistical Computing,” *R Foundation for Statistical Computing (Vienna, Austria).*

Rabin, M. (2000): “Risk Aversion and Expected-Utility Theory: A Calibration Theorem,” *Econometrica*, 68, 1281–1292.

Render, R. A. and X. Walker (1984): “Mixture Densities, Maximum Likelihood and the EM Algorithm,” *SIAM Review*, 26, 195–239.

Starmer, C. (2000): “Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk,” *Journal of Economic Literature*, 38, 332–382.

Stott, H. P. (2006): “Cumulative Prospect Theory’s Functional Menagerie,” *Journal of Risk and Uncertainty*, 32, 101–130.

Tversky, A. and D. Kahneman (1992): “Advances in Prospect Theory: Cumulative Representation of Uncertainty,” *Journal of Risk and Uncertainty*, 5, 297–323.

Wu, G., J. Zhang, and R. Gonzalez (2004): *Decision under Risk*, D. Koehler and N. Harvey (Eds.), The Blackwell Handbook of Judgment and Decision Making, Oxford University Press.

Chapter 3

Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size

This chapter is joint work with Helga Fehr and Thomas Epper.

Another version of this chapter has been published as *SOI Working Paper 0708* and is currently under review at the *Journal of Risk and Uncertainty*.

3.1 Introduction

Risk is a ubiquitous feature of social and economic life. Many of our decisions, such as what trade to learn and where to live, involve risky consequences of great importance. Often these choices entail substantial monetary costs and rewards. Therefore, risk taking behavior under high stakes is a relevant area of economic research. The effect of high stakes on risk tolerance has been debated since the early days of expected utility theory. In a seminal paper, Markowitz (1952) argued that risk preferences are likely to reverse from risk seeking over very small stakes to risk aversion over high stakes. While Markowitz did not test this conjecture experimentally, there is evidence by now that relative risk aversion is indeed greater when substantial amounts of money are at stake (Binswanger, 1981; Kachelmeier and Shehata, 1992; Holt and Laury, 2002), at least on average. Most economists would attribute this change in relative risk aversion to the characteristics of the utility for money, and would search for suitable functional forms that are able to accommodate this behavioral pattern. Little is known about the underlying forces of the increase in relative risk aversion, however. In particular, it is not clear whether this change is actually a consequence of the way people value low versus high amounts of money, or whether some other component of lottery evaluation, such as probability weighting, is the driving force. Moreover, results based on aggregate data may gloss over potentially important differences in individual behavior.

In order to close this gap, we analyze comprehensive choice data stemming from an experiment conducted in Beijing in 2005. The experimental subjects had to take decisions over substantial real monetary stakes with maximum payoffs amounting to more than an average subject's monthly income. The lotteries presented to the subjects were framed as gains and as losses in order to be able to investigate the effect of increasing stake size on relative risk aversion in both decision domains. To disentangle the effects of stake size on the valuation of monetary outcomes and probability weighting, we estimated the parameters of a flexible sign- and rank-dependent decision model, which nests expected utility theory as a special case. Furthermore, to account for the existence of heterogeneous preference types, we used a finite mixture regression model, which assigns each individual to one of several distinct behavioral types and provides type-specific parameter estimates for the underlying decision model (El-Gamal and Grether, 1995; Stahl and Wilson, 1995; Houser et al., 2004).

The following results emerge from our analysis. First, we find a strong and significant framing effect in subjects' evaluations of risky gains and losses. Whereas observed certainty equivalents over rising gains exhibit significantly increasing relative risk aversion, there is no coherent stake-dependent pattern in subject's behavior over identical lotteries framed as losses.

Second, contrary to many economists' expectations, value function parameters remain stable over increased stakes in both decision domains, implying that the observed increase in average relative risk aversion over gains cannot be explained by changing attitudes towards monetary outcomes. Rather, it can be predominantly attributed to a change in probability weighting. The probability weighting function for high stakes deviates less strongly from rational linear weighting than the respective function for low stakes. This change is particularly pronounced over the range of smaller probabilities, entailing less optimistic lottery evaluation at high stakes and, thus, increasing relative risk aversion. In the loss domain, however, no such change in probability weights can be inferred from the data.

Third, when allowing for heterogeneity of preference types, we find two distinct behavioral groups: The majority of about 73% of the subjects exhibit an inverted S-shaped probability weighting curve, whereas the minority can essentially be characterized as expected value maximizers. Furthermore, we show that the observed increase in average relative risk aversion over gains can exclusively be attributed to a change in behavior by the majority group of decision makers, who evaluate high-stake prospects more cautiously by putting lower weights on stated gain probabilities. In contrast, the minority type's behavior is not affected by rising stakes at all.

Our results entail material consequences for decision theory as well as applied economics. The first two findings, the framing effect as well as the probability weighting function as carrier of changing risk attitudes, effectively rule out expected utility theory as a candidate for explaining increasing relative risk aversion at the aggregate level. Since it is the probability weights which are responsible for the change in relative risk aversion, using more flexible utility functions cannot adequately solve the problem of modelling increasing risk aversion. While the observed framing effect lends support to using sign-dependent decision models, such as prospect theory, stake dependence of probability weights, however, calls theories based on stake-invariant probability weighting into question. The third finding poses a challenge to type-independent models of choice under risk, which might be prone to aggregation bias. We show that the vast heterogeneity in individual risk taking behavior, typically found in choice data (Hey and Orme, 1994), is substantive in the sense that a single preference model is unable to adequately describe behavior. This heterogeneity may render policy recommendations based on average parameter estimates inappropriate. Moreover, the mix of different types may be decisive for aggregate outcomes, as the literature on the role of bounded rationality under strategic complementarity and substitutability has shown (Haltiwanger and Waldman, 1985, 1989; Fehr and Tyran, 2005). The researcher will, therefore, need to deal with the potential stake sensitivity of the pivotal group.

To the best of our knowledge, this is the first study that provides a systematic examination of stake effects on probability weights for real substantial payoffs. Neither are we aware of any other study that examines the relevance of framing and type heterogeneity for the impact of stakes on risk tolerance. Previous studies have focused on gains, on hypothetical payoffs or on quite limited payoff ranges, and have usually not addressed the issue of nonlinear probability weighting (Hogarth and Einhorn, 1990; Camerer, 1991; Bosch-Domenech and Silvestre, 1999; Kuehberger et al., 1999; Etchart-Vincent, 2004; Weber and Chapman, 2005). One exception is the experiment by Etchart-Vincent (2004), which investigates the probability weighting function under low and high hypothetical losses. As Holt and Laury (2002, 2005) have convincingly demonstrated, however, it may make a difference whether choices are hypothetical or for real money. They found that, contrary to real payoffs, relative risk aversion does not change significantly with increasing hypothetical stakes.

One of the few previous experiments using substantial real monetary incentives, namely the study by Kachelmeier and Shehata (1992), was also conducted in Beijing. Our work distinguishes itself from theirs in several important respects, however. First of all, our experimental design is not confined to lotteries framed as gains. Second, observed certainty equivalents in our data set show a clearly defined fourfold pattern of risk attitudes for both low stakes and high stakes, i.e. risk aversion for high-probability gains and low-probability losses, as well as risk seeking for low-probability gains and high-probability losses, whereas Kachelmeier and Shehata find practically no risk aversion in choices over low stakes. The authors attribute this lack of risk aversion to the specifics of their elicitation procedure: Certainty equivalents were elicited as minimum selling prices, which seems to have induced a kind of loss aversion in subjects' responses. This might have led otherwise risk averse subjects to reveal risk neutral or risk seeking prices rather than lose their lottery endowment (Kachelmeier and Shehata (1992), p. 1133). What Kachelmeier and Shehata do find in observed certainty equivalents, however, is that stake size interacts significantly with probability level, which is in line with our findings, but their data set is not sufficiently rich to draw any conclusions on relative contributions of outcome valuation and probability weighting, nor do they address the issue of heterogeneity.

The paper is structured as follows. Section 3.2 describes the experimental design and procedures. The decision model applied to the experimental data as well as the finite mixture regression model are presented in Section 3.3. The results of the estimation procedure are discussed in Section 3.4. Section 3.5 concludes the paper.

3.2 Experiment

In the following section, the experimental setup and procedures are described. The experiment took place in Beijing in November 2005. The subjects were recruited by flier distributed at the campusses of Peking University and Tsinhua University. Interested people had to register by email for one of two sessions conducted on the same day. Participants were selected to guarantee a balanced distribution of genders and fields of study. In total, 153 subjects' responses were analyzed.

The experiment served to elicit certainty equivalents for 56 two-outcome lotteries. Twenty-eight lotteries offered low-stake outcomes ranging from 4 to 55 Chinese Yuan

Table 3.1: Gain Lotteries $(x_1, p; x_2)$

p	x_1	x_2	p	x_1	x_2	p	x_1	x_2
0.05	15	4	0.25	250	65	0.75	250	65
0.05	20	7	0.25	320	130	0.75	320	130
0.05	55	20	0.50	7	4	0.90	7	4
0.05	250	65	0.50	15	4	0.90	130	65
0.05	320	130	0.50	20	7	0.95	15	4
0.05	950	320	0.50	130	65	0.95	20	7
0.10	7	4	0.50	250	65	0.95	250	65
0.10	130	65	0.50	320	130	0.95	320	130
0.25	15	4	0.75	15	4			
0.25	20	7	0.75	20	7			

Outcomes x_1 and x_2 are denominated in Chinese Yuan.
p denotes the probability of the higher gain.

(CHN), another 28 lotteries entailed high-stake outcomes from 65 to 950 CHN[1]. Average earnings per subject amounted to approximately 323 CHN, including a show up fee of 20 CHN. Monetary incentives were substantial given the participants' average monthly disposable income of about 700 CHN. Besides, the low-stake outcomes were quite salient by themselves, as the expected payoff over low-stake lotteries amounted to about 16 CHN, considerably more than the going hourly wage rate. Probabilities of the lotteries' higher gain or loss varied from 5% to 95%. One half of the lotteries were framed as choices between risky and certain gains ("gain domain"); the same decisions were also presented as choices between risky and certain losses ("loss domain"). For each lottery in the loss domain, subjects were provided with a specific endowment which served to cover their potential losses. These initial endowments rendered the expected payoff for each loss lottery equal to the expected payoff of an equivalent gain lottery. The set of gain lotteries is presented in Table 3.1.

Subjects were entitled to one random draw from their low-stake decisions and to one random draw from their high-stake decisions. In order to preclude order effects, low-stake and high-stake lotteries were intermixed and appeared in random order in a booklet containing the decision sheets.

For each lottery, a decision sheet, such as presented in Figure 3.1, contained the

[1]At the time of the experiment one Chinese Yuan equalled about 0.12 U.S. Dollars.

Figure 3.1: Design of the Decision Sheet

Decision situation:
22

	Option A	Your Choice:				Option B Guaranteed payoff amounting to:
1		A		o	B	20
2		A		o	B	19
3		A		o	B	18
4		A		o	B	17
5		A		o	B	16
6		A		o	B	15
7		A		o	B	14
8		A	o		B	13
9	Profit of CHN 20 with	A	o		B	12
10	probability 25% and profit of	A	o		B	11
11	CHN 0 with probability 75%	A	o		B	10
12		A	o		B	9
13		A	o		B	8
14		A	o		B	7
15		A	o		B	6
16		A	o		B	5
17		A	o		B	4
18		A	o		B	3
19		A	o		B	2
20		A	o		B	1

specifics of the lottery and a list of 20 equally spaced certain outcomes ranging from the lottery's maximum payoff to the lottery's minimum payoff. Subjects had to indicate whether they preferred the lottery or the certain payoff for each row of the decision sheet. The lottery's certainty equivalent was calculated as the arithmetic mean of the smallest certain amount preferred to the lottery and the following certain amount on the list, when subjects had, for the first time, reported preferring the lottery. For example, if a subject had decided as indicated by the small circles in Figure 3.1, her certainty equivalent would amount to 13.5 CHN.

Before subjects were permitted to start working on the experimental decisions, they were presented with two hypothetical choices to become familiar with the procedure. Subjects could work at their own speed. The vast majority of them needed considerably less than 90 minutes to complete the experiment. At the end of the experiment, one of their low-stake choices and one of their high-stake choices were randomly selected for payment. Subjects were paid in private afterward.

3.3 Econometric Model

This section discusses the specification of the finite mixture regression model, which allows controlling for latent heterogeneity in risk taking behavior. Such an approach requires several building blocks: first, specifying the basic decision model; second, allowing for potentially different behaviors under low and high stakes; third, specifying the error term; and finally, accounting for heterogeneity in behavior.

3.3.1 The Basic Decision Model

The underlying model of decision under risk should be able to accommodate a wide range of different behaviors. Sign- and rank-dependent models, such as cumulative prospect theory (CPT), capture two robust empirical phenomena: nonlinear probability weighting and loss aversion (Starmer, 2000). Therefore, a flexible approach, such as proposed by CPT, lends itself to describing risk taking behavior. According to CPT, an individual values any binary gamble $\mathcal{G}_g = (x_{1g}, p_g; x_{2g}),\ g \in \{1, \ldots, G\}$, where $|x_{1g}| > |x_{2g}|$, by

$$v\left(\mathcal{G}_g\right) = v(x_{1g})w(p_g) + v(x_{2g})(1 - w(p_g)). \tag{3.1}$$

The function $v(x)$ describes how monetary outcomes x are valued, whereas the function $w(p)$ assigns a subjective weight to every outcome probability p. The gamble's certainty equivalent $\hat{ce}_g$ can then be written as

$$\hat{ce}_g = v^{-1}\left[v(x_{1g})w(p_g) + v(x_{2g})(1 - w(p_g))\right]. \tag{3.2}$$

In order to make CPT operational we have to assume specific functional forms for the value function $v(x)$ and the probability weighting function $w(p)$. A natural candidate for $v(x)$ is a sign-dependent power function

$$v(x) = \begin{cases} x^{\alpha} & \text{if } x \geq 0 \\ -(-x)^{\beta} & \text{otherwise,} \end{cases} \tag{3.3}$$

which can be conveniently interpreted and which has also turned out to be the best compromise between parsimony and goodness of fit in the context of prospect theory (Stott, 2006). For this specification of the value function, the existence of loss aversion can be inferred from the difference in the domain-specific curvatures. According to Tversky and Kahneman, loss aversion, in the sense that "losses loom larger than corresponding gains", is present if $v'(x) < v'(-x)$ for $x \geq 0$ (Tversky and Kahneman (1992), p. 303). This is the case if the estimated β is significantly larger than α.

A variety of functions for modeling probability weights $w(p)$ have been proposed in the literature (Quiggin, 1982; Tversky and Kahneman, 1992; Prelec, 1998). We use the two-parameter specification suggested by Goldstein and Einhorn (1987) as well as by Lattimore et al. (1992):

$$w(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma}, \ \delta \geq 0, \ \gamma \geq 0. \tag{3.4}$$

We favor this specification because it has proven to account well for individual heterogeneity (Wu et al., 2004) and its parameters have an intuitively appealing interpretation:[2] The parameter γ largely governs the slope of the curve, whereas the parameter δ largely governs its elevation. The smaller the value of γ, the more strongly the probability weighting function deviates from linear weighting. The larger the value of δ, the more elevated the curve, ceteris paribus. Linear weighting is characterized by $\gamma = \delta = 1$. In a sign-dependent model, the parameters may take on different values for gains and for losses, yielding a total of six behavioral parameters to be estimated.

3.3.2 Stake Dependence

In order to address our focal question of stake size effects, we introduce a dummy variable $HIGH$ into the basic decision model, such that $HIGH = 1$ if the lottery under consideration contains high-stake payoffs amounting to at least 65 CHN, and $HIGH = 0$ otherwise. Each one of the model parameters $\omega \in \{\alpha, \beta, \gamma', \delta'\}$, with γ' and δ' containing the domain-specific parameters for the slope and the elevation of the probability weighting functions, is assumed to depend linearly on $HIGH$ in the following fashion:

$$\omega = \omega_0 + \omega_{HIGH} \times HIGH, \tag{3.5}$$

[2] Moreover, the function generally fits equally well as the two-parameter functional developed by Prelec (1998).

with ω_0 representing the respective low-stake parameters. This step adds another six additional behavioral parameters to be estimated. If relative risk aversion indeed changes with stake size, at least one of the coefficients of the high-stake dummy $HIGH$ should turn out to be significantly different from zero. In particular, if the valuation of monetary outcomes is the driving force behind changing risk tolerance, the respective coefficients should be material in size and statistically significant, since the power functional, used for estimation, cannot account for changing relative risk aversion. If the estimates of α_{HIGH} or β_{HIGH} were indeed significant, the present model would be mis-specified and an alternative specification of the value function that can account for changing relative risk aversion would be called for.

3.3.3 Error Specification

We now turn to the next step of model specification. In the course of the experiment, risk taking behavior of individual $i \in \{1, \ldots, N\}$ was measured by her certainty equivalents ce_{ig} for a set of different lotteries. Since CPT explains *deterministic* choice, the predicted certainty equivalents $\hat{ce}_g$ are bound to deviate from the actual certainty equivalents ce_{ig} by an error ϵ_{ig}, i.e. $ce_{ig} = \hat{ce}_g + \epsilon_{ig}$. There may be different sources of error, such as carelessness, hurry or inattentiveness, resulting in accidentally wrong answers (Hey and Orme, 1994). The Central Limit Theorem supports the assumption that the errors are normally distributed and simply add white noise.

Furthermore, we allow for three different sources of heteroskedasticity in the error variance. First, for each lottery subjects have to consider 20 certain outcomes, which are equally spaced throughout the lottery's range $|x_{1g} - x_{2g}|$. Since the observed certainty equivalent ce_{ig} is calculated as the arithmetic mean of the smallest certain amount preferred to the lottery and the following certain amount, where the lottery is preferred, the error is proportional to the lottery range, which has to be taken account of by the estimation procedure. Second, as subjects may be heterogeneous with respect to their previous knowledge, their ability of finding the correct certainty equivalent as well as their attention span, we expect the error variance to differ by individual. Third, lotteries in the gain domain may be evaluated differently from the ones in the loss domain. Therefore, we allow for domain-specific variance in the error term. This yields the form

$$\sigma_{ig} = \xi_i |x_{1g} - x_{2g}| \tag{3.6}$$

for the standard deviation of the error term distribution, where ξ_i denotes an individual domain-specific parameter. Note that the model allows to test for both individual-specific and domain-specific heteroskedasticity by either imposing the restriction $\xi_i = \xi$, or by forcing all the ξ_i to be equal in both decision domains. Both restrictions are rejected by their corresponding likelihood ratio tests with p-values close to zero. Therefore, we control for all three types of heteroskedasticity in the estimation procedure.

3.3.4 Accounting for Heterogeneity

A suitable estimation procedure, such as maximum likelihood, yields estimates for the average values of the parameters $\theta = (\alpha', \beta', \gamma', \delta')'$. If there is heterogeneity of a substantive kind, i.e. if there are several distinct data generating processes, estimating a single set of parameters is inappropriate and may render misleading results. For this reason, we estimate a finite mixture model in order to account for heterogeneity. The basic idea of the mixture model is assigning an individual's risk-taking choices to one of C different types of behavior, each characterized by a distinct vector of parameters $\theta_c = (\alpha_c', \beta_c', \gamma_c', \delta_c')'$, $c \in \{1, \ldots, C\}$. The estimation procedure yields estimates of the relative sizes of the different groups, the mixing proportions π_c, as well as the group-specific parameters θ_c of the underlying behavioral model. Given our assumptions on the distribution of the error term, the density of type c for the i-th individual can be expressed as

$$f\left(ce_i, \mathcal{G}; \theta_c, \xi_i\right) = \prod_{g=1}^{G} \frac{1}{\sigma_{ig}} \phi\left(\frac{ce_{ig} - \hat{ce}_g}{\sigma_{ig}}\right), \tag{3.7}$$

where $\phi(\cdot)$ denotes the density of the standard normal distribution and ξ_i accounts for individual-specific heteroskedasticity. Since we do not know *a priori* to which group a certain individual belongs to, the relative group sizes π_c are interpreted as probabilities of group membership. Therefore, each individual density of type c has to be weighted by its respective mixing proportion π_c, which is unknown and has to be estimated as well. Taking the sum over the weighted type-specific densities yields the individual's contribution to the model's likelihood function $L(\Psi; ce, \mathcal{G})$. The log likelihood of the finite mixture regression model is then given by

$$\ln L\left(\Psi; ce, \mathcal{G}\right) = \sum_{i=1}^{N} \ln \sum_{c=1}^{C} \pi_c \, f\left(ce_i, \mathcal{G}; \theta_c, \xi_i\right), \tag{3.8}$$

where the vector $\Psi = (\theta_1', \ldots, \theta_C', \pi_1, \ldots, \pi_{C-1}, \xi_1, \ldots, \xi_N)'$ summarizes the parameters to be estimated.

3.3.5 Estimation

In order to deal with the issues of non-linearity and multiple local maxima encountered when maximizing the likelihood of a finite mixture regression model (McLachlan and Peel, 2000), the iterative Expectation Maximization (EM) algorithm is applied (Dempster et al., 1977). This algorithm also provides an additional feature: It calculates, by Bayesian updating in each iteration, an individual's posterior probability τ_{ic} of belonging to group c. These posterior probabilities τ_{ic} represent a particularly valuable result of the estimation procedure. Not only does the procedure endogenously assign each individual to a specific group, but it also provides a method of judging classification quality. If all the τ_{ic} of the final iteration are either close to zero or one, all the individuals are unambiguously assigned to one specific group. The τ_{ic} can be used to calculate a summary measure of ambiguity, such as the Average Normalized Entropy (El-Gamal and Grether, 1995), in order to gauge the extent of dubious assignments. If classification has been successful, a low measure of entropy should be observed.

Furthermore, entropy measures allow the researcher to discriminate between models with differing numbers of types. Since the finite mixture regression model is defined over a pre-specified number of groups, a criterion for assessing the correct number of groups is called for. In the context of mixture models, classical criteria, such as the Akaike Information Criterion AIC or the Bayesian Information Criterion BIC, are not suited for this purpose (Celeux and Soromenho, 1996). Celeux proposes to use entropy criteria based on the posterior probabilities of group assignment τ_{ic} instead, which are shown to perform better than the classical criteria. For example, if entropy increases when the number of different types is raised from two to three, group assignment of the individuals is less reliable, and the model tends to overfit the data. Therefore, the model with two types is to be preferred.

Various problems may be encountered when maximizing the likelihood function of a finite mixture regression model and, therefore, a customized estimation procedure has to be used, which can adequately deal with these problems. Details of the estimation procedure[3] are discussed in Bruhin et al. (2007).

3.4 Results

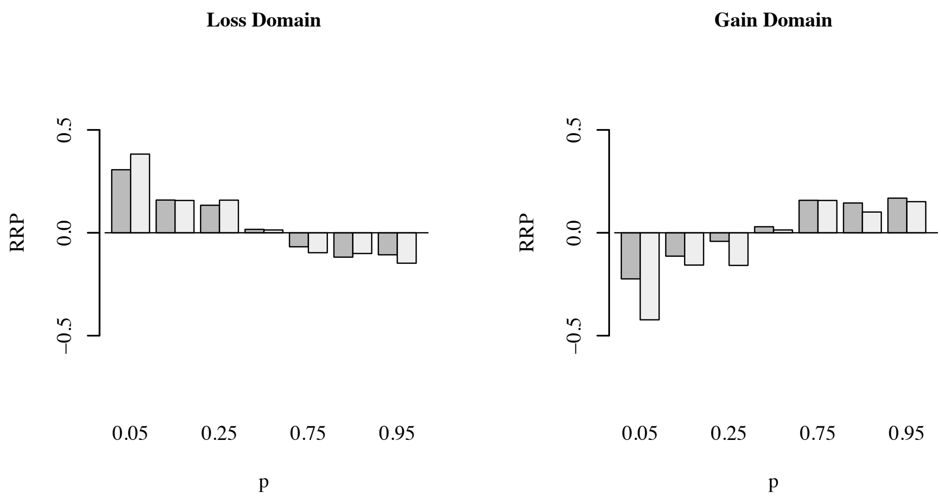
In the following section we investigate the stake-size sensitivity of observed risk taking behavior and present the estimates of the decision model assuming one homogeneous type of preferences. Furthermore, we show that substantive heterogeneity is present in our data and discuss the quality of the classification procedure and the number of heterogeneous behavioral types identified in the data. Finally, we characterize these different types by their average behavioral parameters and discuss the effect of stake size on each group's behavior.

RESULT 1: On average, observed behavior exhibits the fourfold pattern of risk attitudes, predicted by prospect theory, for both low-stake and high-stake outcomes. Stake-specific behavior is subject to a strong framing effect, however: When gains are at stake, relative risk aversion increases with stake size at almost all levels of probability. In the loss domain no such clear picture emerges.

Support. In Figure 3.2, observed risk taking behavior is summarized by the median relative risk premia $RRP = (ev - ce)/|ev|$, where ev denotes the expected value of a lottery's payoff and ce stands for its certainty equivalent. $RRP > 0$ indicates risk aversion, $RRP < 0$ risk seeking, and $RRP = 0$ risk neutrality. The light gray bars in Figure 3.2 represent the observed median RRP for low-stake lotteries, the ones in dark gray represent the respective high-stake median RRP. The median relative risk premia RRP, sorted by the probability p of the higher gain or loss, show a systematic relationship with p: For both low stakes and high stakes, subjects' choices display a fourfold pattern, i.e. they are

[3] The procedure is written in the R environment (R Development Core Team, 2006).

Figure 3.2: Median Relative Risk Premia by Stake Size



Low stakes: light gray. High stakes: dark gray.

risk averse for low-probability losses and high-probability gains, and they are risk seeking for low-probability gains and high-probability losses. Therefore, at first glance, average behavior is adequately described by a model such as CPT.

What the bar plots also reveal is that median relative risk premia differ substantially by stake level: When subjects' preferences exhibit increasing relative risk aversion, we should observe different low-stake and high-stake RRP, namely, high-stake choices should be relatively less risk tolerant than low-stake choices. Inspection of Figure 3.2 confirms that, in the gain domain, median high-stake choices are considerably less risk seeking for small probabilities and somewhat more risk averse for large probabilities than their median low-stake counterparts. For losses, the evidence is not so clear-cut, however. At some levels of probability, low-stake median RRP display relatively higher risk aversion than high-stake RRP, at some other levels the reverse is true.

In order to judge whether the distributions of the stake-dependent RRP are significantly different from each other, we performed a series of Wilcoxon signed-rank tests for each level of probability, which yield the following results at conventional levels of significance: With the exception of the probability of 95%, all the low-stake RRP over gains are significantly smaller than the high-stake ones. We therefore conclude that there is a significant stake effect in the data on choices over gains: On average, people are relatively more risk averse for high gains than for low gains. In the loss domain, no consistent picture emerges: Low-stake RRP are significantly smaller at three levels of probability ($p \in \{0.10, 0.75, 0.95\}$), significantly larger at one level ($p = 0.05$), and insignificantly different at the remaining three levels of probability ($p \in \{0.25, 0.50, 0.90\}$). Therefore, we conclude that there is no obvious systematic relationship between stake-size effect and level of probability for loss lotteries.

Our data show behavior consistent with nonlinear probability weighting, but also a substantial framing effect. Relative risk aversion increases with stake size, albeit only for gains. When subjects evaluate the same lotteries framed as losses rather than as gains, their relative risk aversion does not systematically increase. In fact, no coherent pattern of stake-dependent behavior under losses emerges. This sensitivity to framing, already visible at the level of observed behavior, clearly excludes expected utility theory from the list of eligible models for describing average risk taking behavior.

We now turn to one of our major concerns, namely, whether the change in relative risk aversion over gains can be attributed to a specific component of lottery evaluation. First, we focus on the results for the average parameter estimates without accounting for heterogeneity.

RESULT 2: In the homogeneous preference model, the estimated curvatures of the value functions do not significantly change with rising stakes.

Support. Table 3.2 contains the parameter estimates for the decision model discussed in Section 3.3. For the time being, we focus on the average parameter estimates displayed in the two columns labeled "Pooled". The curvature parameters of the value functions over low stakes are denoted by α_0 for gains and β_0 for losses. α_{HIGH} and β_{HIGH} represent the corresponding estimated coefficients of the high-stake dummy $HIGH$, measuring the change in curvature brought about by increased stake levels. For both domains, the estimates for α_{HIGH} and β_{HIGH} are small in size, and the bootstrapped standard errors, reported in parentheses below the respective point estimates, indicate that the coefficients are not significantly different from zero. Furthermore, when a restricted model with stake-invariant curvature parameters is estimated, the likelihood ratio test of the restricted

Table 3.2: Classification of Behavior

	Gains				Losses		
	Pooled	EUT	Non-EUT		Pooled	EUT	Non-EUT
π		0.266	0.734	π		0.266	0.734
		(0.026)	(0.026)			(0.026)	(0.026)
α_0	0.467	0.996	0.430	β_0	1.165	1.157	1.177
	(0.109)	(0.136)	(0.116)		(0.110)	(0.136)	(0.120)
α_{HIGH}	0.047	-0.080	0.066	β_{HIGH}	-0.038	-0.137	-0.106
	(0.158)	(0.165)	(0.167)		(0.162)	(0.178)	(0.178)
γ_0	0.316	0.863	0.225	γ_0	0.383	0.802	0.284
	(0.012)	(0.067)	(0.013)		(0.012)	(0.067)	(0.012)
γ_{HIGH}	0.056	0.026	0.058	γ_{HIGH}	0.045	0.027	0.046
	(0.012)	(0.023)	(0.012)		(0.012)	(0.024)	(0.012)
δ_0	1.304	0.952	1.265	δ_0	0.913	0.912	0.917
	(0.076)	(0.094)	(0.080)		(0.052)	(0.090)	(0.058)
δ_{HIGH}	-0.324	-0.040	-0.344	δ_{HIGH}	0.070	0.106	0.099
	(0.095)	(0.090)	(0.098)		(0.077)	(0.091)	(0.087)
$\ln L$	31,536	32,580					
Parameters	318	331					
Observations	8,560	8,560					

Standard errors in parentheses are based on the bootstrap with 2,000 replications.
Parameter vectors include estimates of $\hat{\xi}_i$ for domain- and individual-specific error variances.

model against the unrestricted one renders a p-value of 0.911. This test result implies that the hypothesis of equal curvatures over both ranges of outcomes cannot be rejected.

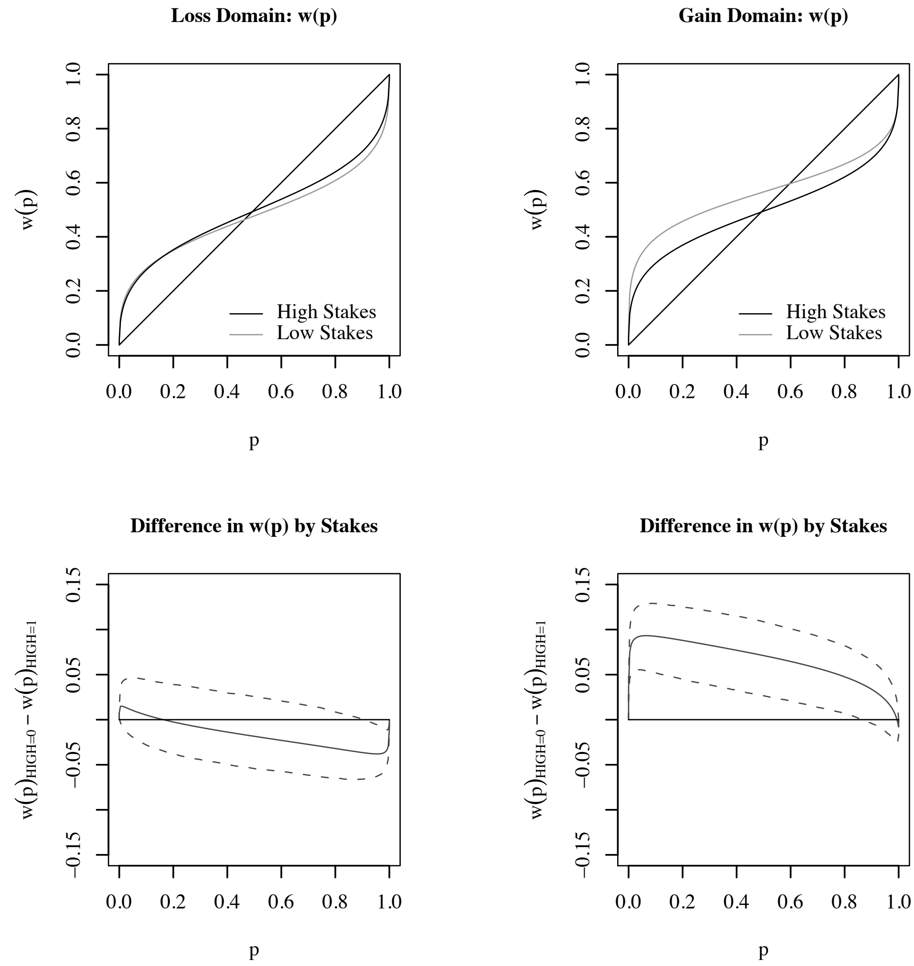
If the valuation of monetary outcomes were the carrier of increasing relative risk aversion over gains, the estimates of α_{HIGH} would have to be statistically significant and, presumably, quite sizable, since the specification of the value function as a power function can only accommodate constant relative risk aversion. As the estimation results show, however, this is not the case. Therefore, we conclude that changing attitudes towards monetary outcomes are not responsible for the observed increase in relative risk aversion. This finding is also holds for alternative specifications of the value function that are sufficiently flexible to capture changing relative risk aversion, such as the expo-power function introduced by Saha (1993).

As the curvature of the value function is robust to stake size, the observed increase in relative risk aversion over gains has to be driven by the other component of lottery evaluation, probability weighting, as the next result confirms.

RESULT 3: For homogeneous preferences, low-gain probability weights deviate more strongly from rational linear weighting than high-gain probability weights. No substantial change in probability weights is observed for losses.

Support. We first discuss the results for the gain domain. A first indication of probability weights being the carrier of the observed stake effect for gains can already be found in the bar plots in Figure 3.2. The *differences* in the stake-dependent observed RRP decrease markedly with increasing probability level, suggesting a substantial interaction effect. Inspection of the "Pooled"-column in the gains section of Table 3.2 indeed confirms that the estimated change in the elevation of the curve, measured by δ_{HIGH}, is significantly negative and substantial in size, implying a major decrease in elevation from 1.307 to 0.979, induced by substantially less optimistic weighting of probabilities. Moreover, the change in the slope of the probability weighting function γ_{HIGH} is significantly positive (0.056), implying a slightly less strongly S-shaped curve for high stakes. The impact of these parameter changes on the shape of the probability weighting function can be examined in Figure 3.3. The top panel of the figure shows, for each decision domain, the estimated probability weighting curves for low stakes, defined by $HIGH = 0$, plotted

Figure 3.3: Average Probability Weights by Stake Size



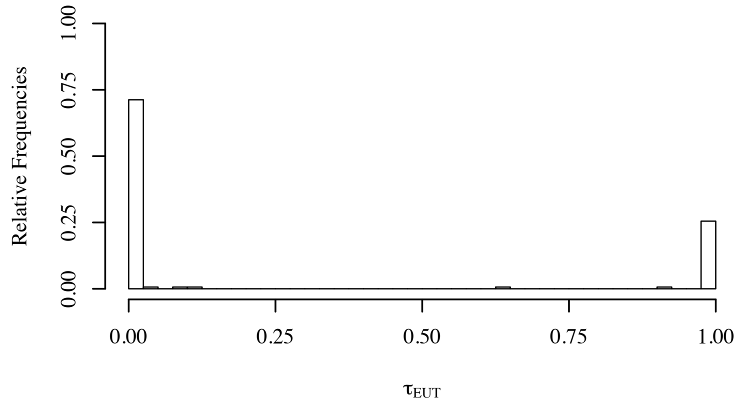
Dashed lines: 95%-confidence bands based on the percentile bootstrap method against the high-stake curves, defined by $HIGH = 1$. Evidently, the low-gain function is much more elevated and slightly more strongly curved than the high-gain function.

However, significant changes in single parameter estimates do not tell the whole story. Since the probability weights are a nonlinear combination of two parameters, inference needs to be based on γ and δ jointly. Therefore, the percentile bootstrap method (using 2,000 replications) was employed to construct the 95%-confidence bands for the difference in the low-stake and the high-stake probability weighting curves. In order to judge the overall effect of rising stakes on the shape of the probability weighting function, we inspect the bottom panel of Figure 3.3, depicting the confidence bands for the stake-dependent differences in probability weights. Whenever a confidence band includes the zero line, the hypothesis of stake-invariant probability weights cannot be rejected. The graph on the right hand side for the gain domain, however, shows that the difference between low-gain probability weights and high-gain probability weights is indeed statistically significant practically over the whole range of probabilities. Therefore, we have conclusive evidence that the high-gain probability weighting curve departs significantly and substantially from the low-gain curve.

These findings demonstrate that probability weights are the carrier of changing risk tolerance, and suggest that prospect theory, and for that matter many other decision theories which postulate stake-independent probability weighting, cannot adequately deal with risk taking choices involving major changes in stake levels.

In the domain of losses, a totally different picture emerges. The left hand side of the top panel of Figure 3.3 depicts practically overlapping low-loss and high-loss probability weighting curves. The high-loss curve is slightly less strongly S-shaped, which is also reflected in the significant parameter estimate for γ_{HIGH}, amounting to 0.045 (Table 3.2). However, this immaterial difference in the stake-dependent slope parameters does not imply a significant difference in the overall shape of the curves: The bottom panel of Figure 3.3 shows that the 95%-condidence band for the difference in the stake-dependent probability weighting curves includes the zero line practically for all levels of probability. This finding implies that, in choices framed as losses, stake effects are negligible, in line with the lack of any stake-dependent pattern diagnosed in the observed RRP.

Figure 3.4: Posterior Probabilities of Being an EUT Type



So far we have only considered the evidence for the average decision maker. If there is heterogeneity in the population, in the sense that a single preference theory cannot adequately capture behavior, the parameter estimates of the pooled model may be misleading. For this reason, the analysis is extended to account for latent heterogeneity by estimating a finite mixture regression model.

The first question to be answered concerns the number of different types present in the population. One way of dealing with this question is calculating a measure of entropy for varying numbers of groups and choosing the model with the lowest entropy.

RESULT 4: There is substantive heterogeneity in individual risk preferences, which can be captured by two distinct types of behavior. Assuming three distinct types yields an inferior characterization of the underlying heterogeneity.

Support. The finite mixture regression model classifies individuals according to a given number of types. In order to evaluate the quality of classification, we calculated the Average Normalized Entropy ANE (El-Gamal and Grether, 1995) defined as

$$ANE = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\tau_{ic}\log_C\left(\tau_{ic}\right), \tag{3.9}$$

for C groups and N individuals. Taking $\log_C$ normalizes the entropy measure to lie within $[0,1]$. If all the probabilities of individual group membership τ_{ic} are equal to zero or one,

$ANE = 0$. In this case, all the individuals can be perfectly assigned to one group. ANE $= 1$ reflects maximum entropy, i.e. all the τ_{ic} are equal to $1/C$. Such a result indicates that group membership is totally ambiguous and that categorization has failed. For two groups, we find an ANE-value of 1.8% of the maximum entropy. Given this extremely low degree of ambiguity in our two-group classification, an improvement in entropy, when three groups are assumed, seems hardly possible. If the classification procedure worked better for three groups than for two groups, the average normalized entropy should be smaller for $C = 3$ than for $C = 2$. This is clearly not the case, as ANE amounts to 3.2% for the three-group classification. So we can safely conclude that two types of behavior are sufficient to capture the essential characteristics of individual heterogeneity in risk taking.

The low value of ANE in our analysis indicate that nearly all the individuals can be unambiguously assigned to one of the two types. This clean segregation can also be inferred from the distributions of the posterior probabilities of group assignment in Figure 3.4: τ_{EUT} denotes the posterior probability of belonging to the first group, which can be characterized, as we will demonstrate below, as expected utility maximizers ("EUT types"): The individuals' posterior probabilities of being an expected utility maximizer are either close to one or close to zero for practically all the individuals. The histogram also shows that the EUT group encompasses a minority of the decision makers, whereas the other group represents a majority of close to 75% of the subjects.

The subsequent group of results addresses the focal questions: How can these two different types be characterized? And in which way do they react to rising stake levels?

RESULT 5: The minority type, constituting about 27% of the subjects, can essentially be characterized by expected value maximization over both low- and high-outcome ranges.

Support. There are several pieces of evidence in support of *RESULT 5.* Let us first turn to Table 3.2, which also contains the parameter estimates for the finite mixture regression model. The estimates for the minority type are displayed in the columns labeled "EUT". The relative group size of the minority type is estimated to be 0.266, matching the size of the corresponding bar in the histogram of Figure 3.4.

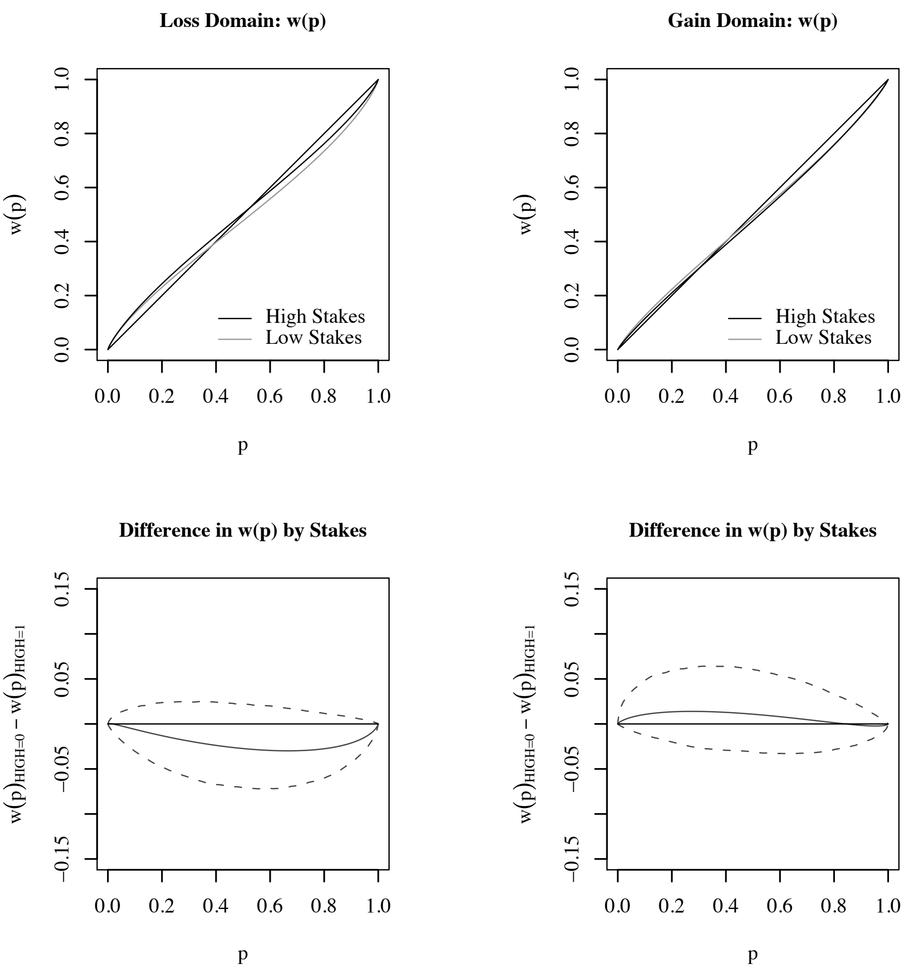
In order to be able to characterize decisions as consistent with expected value maximization, both the value functions and the probability weighting functions are required to be linear. Turning to outcome valuation, we observe that α_0 and β_0 are not statistically distinguishable from one, as the standard errors in Table 3.2 reveal. Furthermore, the coefficients of the high-stake dummy are insignificantly different from zero, indicating the robustness of the value function curvatures to increasing stake size. Therefore, we conclude that the value functions over both gains and losses are essentially linear and unresponsive to stake size.

Linearity of the second model component, probability weighting, holds if the parameter estimates for both γ and δ are equal to one. In Table 3.2 the low-stake parameter estimates for δ_0 are not distinguishable from one, but the respective ones for γ_0 are. However, inspection of the probability weighting curves in Figure 3.5 confirms that departures from linear probability weighting are insubstantial. Furthermore, for both gains and losses, no stake-size effect is visible in slope nor elevation of the probability weighting curves, as both γ_{HIGH} and δ_{HIGH} are insignificantly different from zero, and the 95%-confidence bands for the difference in the stake-dependent probability weighting curves include the zero line, as confirmed by the bottom panel of Figure 3.5. These findings suggest that the first type of decision makers behaves essentially as expected value maximizers, and therefore consistently with EUT. These conclusions, based on the estimation results, also bear out at the level of observed behavior. The EUT types' median relative risk premia in the bottom panel of Figure 3.7 are close to zero, indicating near risk neutrality for both low stakes and high stakes.

Obviously, this first group's, the minority's, behavior is robust over the whole outcome range and can, therefore, not account for increasing relative risk aversion observed in the aggregate data. As the next result shows, the second group of individuals, constituting approximately 73% of the subjects, exhibit a completely different set of behavioral parameter values.

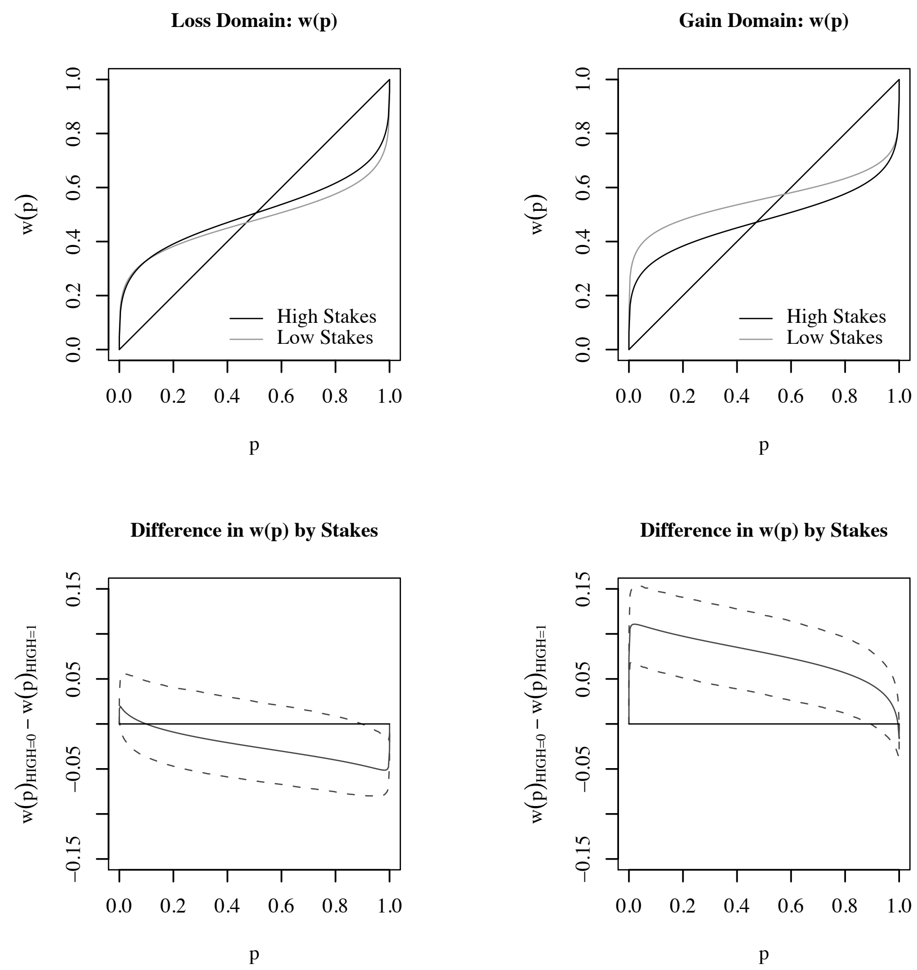
RESULT 6: The majority, Non-EUT types', behavior is characterized by nonlinear probability weighting. Whereas value function parameters remain stable over the whole outcome range in both decision domains, probability weights for gains do not. The low-gain probability weighting curve is characterized by a significantly more optimistic weighting of

Figure 3.5: Probability Weights by Stake Size of EUT Types



Dashed lines: 95%-confidence bands based on the percentile bootstrap method.

Figure 3.6: Probability Weights by Stake Size of Non-EUT Types



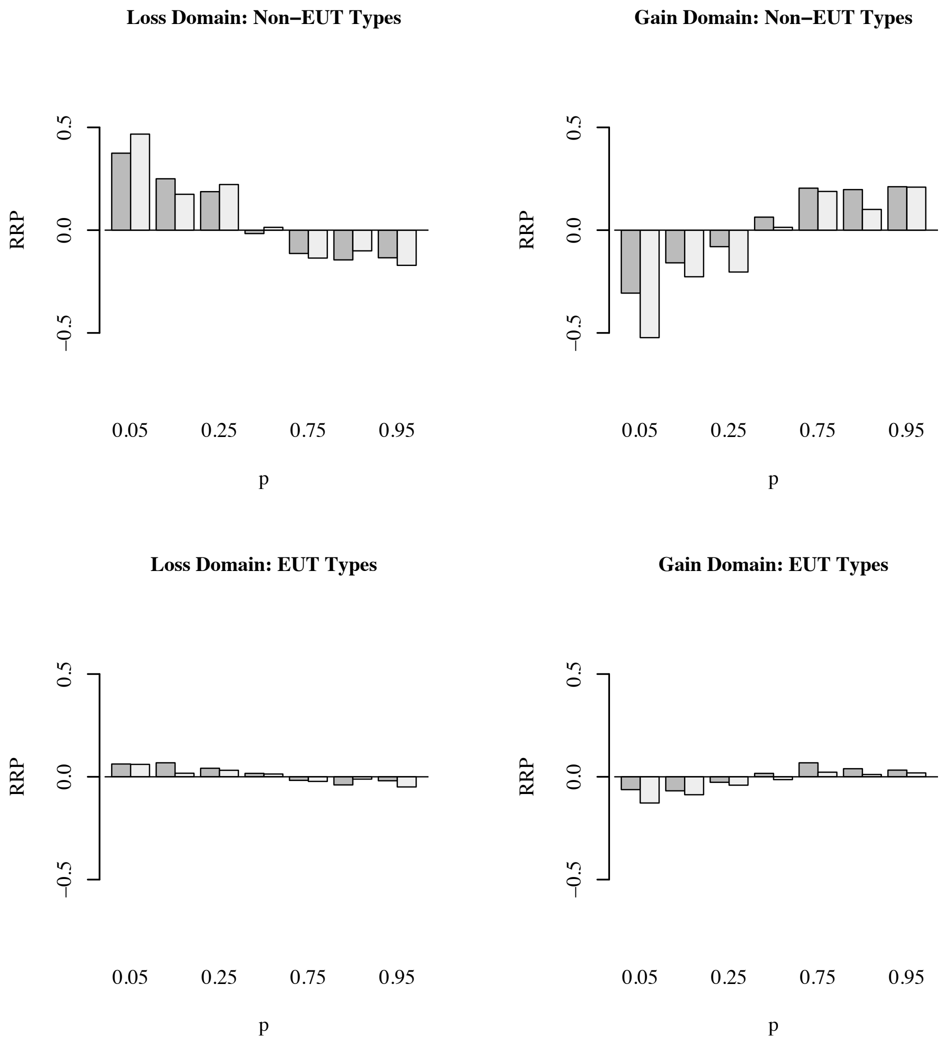
Dashed lines: 95%-confidence bands based on the percentile bootstrap method.

probabilities than the high-gain curve. No such effect is present in the probability weighting curves for losses, however.

Support. The top panel of Figure 3.6 displays the stake-specific probability weighting curves for the second group of individuals labeled "Non-EUT types". For both gains and losses these curves are inverted S-shaped, but there is a major domain-specific difference. In the loss domain the stake-specific curves practically coincide, and their difference is not statistically significant, as the left hand side of the bottom panel of Figure 3.6 confirms. In the gain domain, however, we find the high-stake probability weighting curve to be substantially less elevated than the low-stake one. This change is brought about by the significant stake sensitivity of the elevation parameter over gains, reflected in the corresponding estimate for δ_{HIGH}, which amounts to -0.344 (see the columns labeled "Non-EUT" in Table 3.2). The high-gain probability weighting function is also slightly less curved than the low-gain one, as γ_{HIGH} is estimated to be 0.058. The joint impact of these parameter changes is also statistically significant, as the right hand side of the bottom panel of Figure 3.6 shows. Therefore, we conclude that increasing relative risk aversion over gains is mainly attributable to the Non-EUT types' behavior who weight high-gain probabilities significantly and substantially less optimistically than low-gain ones. These effects can also be traced back in the pattern of observed choices: The top panel of Figure 3.6 displays a substantial stake-dependent difference, particularly over smaller probabilities, in the Non-EUT types' median RRP, which is much more pronounced than the respective difference in the pooled data shown in Figure 3.2.

The results of the finite mixture regression demonstrate that there is substantive heterogeneity in risk taking behavior, which may be glossed over when focusing on a single-preference model. Only one distinct group of individuals is prone to changing risk tolerance when stakes are increased. These Non-EUT types tend to evaluate low-gain prospects significantly more optimistically than high-gain prospects. Thus, prospect theory, even though designed to explain non-EUT behavior, cannot account for this change in relative risk aversion.

Figure 3.7: Type-Specific Median Relative Risk Premia by Stake Size



Low stakes: light gray. High stakes: dark gray.

3.5 Conclusions

This paper pursues three goals. First, it studies the effect of substantial real gains and losses on risk taking. Second, the paper analyzes the influence of rising stakes on the components of lottery evaluation, i.e. on the value and probability weighting functions. Third, it examines heterogeneity in risk taking behavior over varying stake levels. The results of this investigation can be summarized as follows: We find a significant and sizable increase in relative risk aversion when gains are scaled up. In the domain of losses, however, no such clear effect is present in the data. Since subjects evaluate lotteries differently depending on the lotteries being framed as gains or as losses, expected utility theory is effectively ruled out as a valid description of behavior.

Contrary to previous attempts at explaining the increase in relative risk aversion over gains by changing attitudes toward monetary payoffs, the increase can be mainly attributed to a move of the average probability weighting function towards rational linear weighting. As the finite mixture regression analysis shows, this average effect is brought about by the behavior of a majority of decision makers who tend to weight probabilities of low-stake gains considerably more optimistically than probabilities of high-stake gains. Whereas these Non-EUT types' behavior is sensitive to payoff levels, the minority group's behavior, which is shown to be largely consistent with expected value maximization, is not.

These results pose a number of potential problems to both theoretical and applied economics. As most theories of decision under risk typically assume separability of probability weights and outcome valuation, decision models may misrepresent risk preferences considerably when probability weights interact with payoffs. Our results suggest that this interaction effect is significant and substantial in the gain domain, which renders rank-dependent models, such as prospect theory, questionable.

In the field of applied economics, one of the most important issues concerns the substantive heterogeneity found in the population. This study has demonstrated that there are two distinct behavioral types who either weight probabilities near linearly or nonlinearly. A similar distribution of clearly segregated types was also found in two independent Swiss data sets (Bruhin et al., 2007) and, for choices over gains only, in a British data set

(Conte et al., 2007), which suggests that this mix of preference types seems to be quite robust across times and cultures. This substantive kind of heterogeneity has to be taken into account when predicting behavior, as average parameter estimates may be quite misleading. Moreover, as the literature on the role of bounded rationality under strategic complementarity and substitutability has shown, the mix of rational and irrational actors may be crucial for aggregate outcomes (Haltiwanger and Waldman, 1985, 1989; Fehr and Tyran, 2005). This literature demonstrates that, depending on the nature of strategic interdependence, even a minority of players of a particular type may be decisive.

The researcher will, therefore, have to determine which one of the distinct behavioral groups identified in the population will most likely dominate aggregate outcomes. If she can safely conclude that the minority EUT type will dominate, stake dependence of risk preferences is not an issue and expected value maximization may be the adequate model of decision making. If, however, she regards the Non-EUT types as decisive for aggregate outcomes, stake dependence might be a serious problem when gains are concerned. If actual stakes are much larger than the ones used for model estimation, predicting behavior on the basis of estimated model parameters might lead to a significant overestimation of risk tolerance. In order to get a handle on choices under substantial stakes, research will probably have to turn to field data to generate meaningful parameter estimates. Since our results suggest that stake-sensitivity is largely due to a change in probability weights, using more flexible specifications of utility functions, as proposed by Holt and Laury (2002), cannot adequately solve the problem of modelling risk preferences.

The issue of stake sensitivity is an important one for choices over gains. When evaluating potential losses, however, subjects seem to use different heuristics and decision rules than when evaluating gains, which renders them rather insensitive to stake size. This stability of behavior can be seen in both the EUT and the Non-EUT groups. While stake dependence is not an issue here, the researcher will still have to worry about heterogeneity and the ensuing type-dominance question. As Mason et al. (2005) have pointed out, the common assumption of linear probability weights may lead to problematic policy recommendations, if nonlinear probability weighting is the dominant pattern of behavior. The results of this paper cast doubt on research strategies that do not take framing effects, nonlinear probability weighting and the substantive heterogeneity of decision makers into account.

Bibliography

Binswanger, H. P. (1981): "Attitudes toward Risk: Theoretical Implications of an Experiment in Rural India," *Economic Journal*, 91, 867–890.

Bosch-Domenech, A. and J. Silvestre (1999): "Does Risk Aversion or Attraction Depend on Income? An Experiment," *Economics Letters*, 65, 265–273.

Bruhin, A., H. Fehr-Duda, and T. F. Epper (2007): "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion," *Socioeconomic Institute, University of Zurich, Working Paper*, 0705.

Camerer, C. (1991): *Recent Tests of Generalizations of Expected Utility Theory*, W. Edwards (ed.), Utility: Measurement, Theory and Applications. Amsterdam, Kluwer, Nijhoff.

Celeux, G. and G. Soromenho (1996): "An Entropy Criterion for Assessing the Number of Clusters in a Mixture Model," *Journal of Classification*, 13, 195–212.

Conte, A., J. D. Hey, and P. G. Moffat (2007): "Mixture Models of Choice Under Risk," *Department of Economics and Related Studies, University of York, Discussion Paper*, 2007/06.

Dempster, A., N. Laird, and D. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

El-Gamal, M. A. and D. M. Grether (1995): "Are People Bayesian? Uncovering Behavioral Strategies," *Journal of the American Statistical Association*, 90, 1137–1145.

Etchart-Vincent, N. (2004): "Is Probability Weighting Sensitive to the Magnitude of Consequences? An Experimental Investigation on Losses," *Journal of Risk and Uncertainty*, 28, 217–235.

Fehr, E. and J. R. Tyran (2005): "Individual Irrationality and Aggregate Outcomes," *Journal of Economic Perspectives*, 19, 43–66.

Goldstein, W. and H. Einhorn (1987): "Expression Theory and the Preference Reversal Phenomena," *Psychological Review*, 94, 236–254.

Haltiwanger, J. C. and M. Waldman (1985): "Rational Expectations and the Limits of Rationality: An Analysis of Heterogeneity," *American Economic Review*, 75, 326–340.

——— (1989): "Limited Rationality and Strategic Complements: The Implications for Macroeconomics," *Quarterly Journal of Economics*, 104, 463–483.

Hey, J. D. and C. Orme (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62, 1291–1326.

Hogarth, R. M. and H. J. Einhorn (1990): "Venture Theory: A Model of Decision Weights," *Management Science*, 36, 780–803.

Holt, C. A. and S. K. Laury (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655.

——— (2005): "Risk Aversion and Incentive Effects: New Data without Order Effects," *American Economic Review*, 95, 902–904.

Houser, D., M. Keane, and K. McCabe (2004): "Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence using a Bayesian Type Classification Algorithm," *Econometrica*, 72, 781–822.

Kachelmeier, S. J. and M. Shehata (1992): "Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence from the People's Republic of China," *American Economic Review*, 82, 1120–1141.

Kuehberger, A., M. Schulte-Mecklenbeck, and J. Perner (1999): "The Effects of Framing, Reflection, Probability, and Payoff on Risk Preference in Choice Tasks," *Organizational Behavior and Human Decision Processes*, 78, 204–231.

Lattimore, P. K., J. R. Baker, and A. D. Witte (1992): "The Influence of Probability on Risky Choice," *Journal of Economic Behavior and Organization*, 17, 377–400.

Markowitz, H. (1952): "The Utility of Wealth," *The Journal of Political Economy*, 60, 151–158.

Mason, C. F., J. F. Shogren, C. Settle, and J. A. List (2005): "Investigating Risky choices Over Losses Using Experimental Data," *Journal of Risk and Uncertainty*, 31, 187–215.

McLachlan, G. and D. Peel (2000): *Finite Mixture Models*, Wiley Series in Probabilities and Statistics.

Prelec, D. (1998): "The Probability Weighting Function," *Econometrica*, 3, 497–527.

Quiggin, J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3, 323–343.

R Development Core Team (2006): "R: A Language and Environment for Statistical Computing," *R Foundation for Statistical Computing (Vienna, Austria).*

Saha, A. (1993): "Expo-Power Utility: A 'Flexible' Form for Absolute and Relative Risk Aversion," *American Journal of Agricultural Economics*, 75, 905–913.

Stahl, D. O. and P. W. Wilson (1995): "On Players' Models of other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10, 218–254.

Starmer, C. (2000): "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk," *Journal of Economic Literature*, 38, 332–382.

Stott, H. P. (2006): "Cumulative Prospect Theory's Functional Menagerie," *Journal of Risk and Uncertainty*, 32, 101–130.

Tversky, A. and D. Kahneman (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

Weber, B. J. and G. B. Chapman (2005): "Playing for Peanuts: Why is Risk Seeking More Common for Low-Stakes Gambles?" *Organizational Behavior and Human Decision Processes*, 97, 31–46.

Wu, G., J. Zhang, and R. Gonzalez (2004): *Decision under Risk*, D. Koehler and N. Harvey (Eds.), The Blackwell Handbook of Judgment and Decision Making, Oxford University Press.

Chapter 4

Stochastic Expected Utility and Prospect Theory in a Horse Race: A Finite Mixture Approach

Another version of this chapter has been published as *SOI Working Paper 0803*.

4.1 Introduction

Many economic decisions, especially the most important ones, such as choosing the optimal career, asset allocation, or partner, involve risky consequences. Even though a sound understanding of how individuals deal with uncertain outcomes is crucial for characterizing various markets' outcomes there is, so far, no single best model for individual decision making under risk. To explain the St. Petersburg Paradox, Daniel Bernoulli hypothesized in 1738 that individuals maximize their expected utility, which is computed as the sum of utilities of a lottery's outcomes weighted by their corresponding probabilities of realization (Bernoulli, 1954). Expected utility became a cornerstone of standard microeconomic theory as it applies to any regular preference relation defined over a finite number of states (von Neumann and Morgenstern, 1947). However, there is abundant empirical evidence indicating that expected utility theory in its standard form is violated (Starmer, 2000). For example, individuals tend to be risk seeking for small-probability gains and large-probability losses, whereas they are risk averse for large-probability gains and small-probability losses. This fourfold pattern of risk attitudes (Tversky and Kahneman, 1992), where individuals switch between risk averse and risk seeking behavior depending on the outcomes' probabilities, is incompatible with expected utility maximization. In light of these descriptive shortcomings of expected utility theory a plethora of alternative decision models have been developed (Starmer, 2000).

The most prominent alternative is Prospect Theory (PT) (Kahneman and Tversky, 1979), which offers a psychologically plausible account of expected utility theory violations, based on the notion of diminishing sensitivity. In PT individuals evaluate prospects with respect to a specific reference point which defines monetary outcomes as gains or as losses. The value functions over gains and over losses are both characterized by declining rates of marginal value and, thus, result in a typical S-shaped curve. As certainty and impossibility constitute obvious reference points as well, any deviations from probabilities of either zero or one are perceived at a diminishing rate of sensitivity, which leads to characteristically inversely S-shaped probability weighting functions. Such a tendency to overweight small and underweight large probabilities, in conjunction with the sign-dependent valuation of monetary outcomes, directly implies a fourfold pattern of risk attitudes. Consequently, PT and its rank-dependent extension to Cumulative Prospect Theory (Tversky and Kahneman, 1992) turn out as some of the best fitting models for aggregate choices (Hey and Orme, 1994; Camerer and Ho, 1994). Besides its descriptive qualities, recent studies in neuroeconomics and evolutionary psychology indicate that PT even seems to have a neuronal representation in the frontal regions of the brain (Trepel et al., 2005; Camerer et al., 2005), the origins of which may be explained by optimal foraging theory (McDermott et al., 2008). To achieve a good fit, however, parametric models based on PT tend to require rather complex specifications and large data sets, which makes estimation at the individual level often difficult. Moreover, PT is silent on the determinants of the reference point for evaluating monetary outcomes.

Another problem common to all deterministic decision models, such as expected utility theory and PT, is their inability to describe preference instability. Various studies report subjects to reverse their preferences in roughly 25%-45% of the cases when facing the same decisions for a second time (Camerer, 1989; Starmer and Sudgen, 1989; Wu, 1994; Hey and Orme, 1994). Since such a behavior contradicts deterministic choice, researchers often introduce some kind of ad-hoc stochastic error to make their models operational. Blavatskyy (2007), on the other hand, develops a more elaborate structure for the error term which constitutes Stochastic Expected Utility Theory's (SEUT) key feature.

In SEUT, individuals behave as expected utility maximizers but make errors when computing a lottery's expected utility. By assumption, the value attributed to any given lottery can never exceed the value of its highest payoff, nor can it fall below the value of its lowest payoff. Since a lottery's most extreme payoffs represent obvious bounds for its valuation, such an assumption not only seems plausible but also is supported by findings of Gneezy et al. (2006) who attribute observed certainty equivalents lying outside the lottery's range solely to errors individuals make when converting payoffs from one denomination to another. Consequently, SEUT implies a truncated error term with a support confined to the lottery's range. Such a truncated error distribution, which is generally asymmetric, directly incorporates the fourfold pattern of risk attitudes, as a lottery's expected utility is likely to get overvalued (undervalued) when it is close to the utility of the lowest (highest) outcome. Since the fourfold pattern in risk taking behavior results from the error structure, SEUT is only a descriptive model and does not explain why empirical violations of expected utility theory come about. Nevertheless, the model remains fairly parsimonious and, as a moderate extension of expected utility theory, fits well into standard microeconomic theory. When comparing SEUT and PT in various well-known data sets, Blavatskyy (2007) attests SEUT a superior performance at describing representative choices. However, these comparisons ignore potential individual heterogeneity and are based on fairly homogeneous subject pools which all stem from developed Western countries.

Furthermore, there is vast heterogeneity in individual risk taking behavior (Hey and Orme, 1994) rendering purely representative agent approaches questionable, especially when markets are imperfect and there is risk of aggregation bias (Fehr and Tyran, 2005). With the advent of finite mixture models in the field (Stahl and Wilson, 1995; El-Gamal and Grether, 1995; Houser et al., 2004), experimental economists are now equipped with a convenient econometric tool to deal with latent individual heterogeneity in a parsimonious way. These models allow identifying and characterizing different behavioral types in the population and provide an endogenous individual classification into these types. Independent studies by Conte et al. (2007) and Bruhin et al. (2007) apply finite mixture specifications to a total of four different experimental data sets on risk taking behavior and find roughly 20% of the participants to behave essentially risk neutrally, whereas the majority of about 80% of the participants clearly exhibit the fourfold pattern of risk attitudes.

For two quite diverse populations, this study first examines PT's and SEUT's performance in fitting aggregate choices. It uses data from two different experiments which were conducted in Zurich, Switzerland as well as in Beijing, People's Republic of China. In both experiments, which have the same basic design in common, the certainty equivalents of a large number of binary lotteries, framed either as gains or losses, are elicited for a total of 271 participants. In contrast to Blavatskyy (2007), the results on the aggregate level are mixed since, depending on the data set, either PT or SEUT superiorly describe a representative agent's choices. In fact, an inspection of the individual mean squared errors reveals that SEUT provides a superior fit compared to PT for only roughly one third of the participants in both data sets. Such stable shares call representative agent approaches into question and suggest a mix of theories, as applied in the second part of the analysis.

To control for individual heterogeneity, a finite mixture model estimates the behavioral parameters of two types, one PT the other SEUT, while it endogenously determines which one of the two theories best describes a specific subject's choices. In both data sets the resulting individual classifications are remarkably clean and robust: With low measures of entropy, about 25% of the subjects are assigned to the SEUT group whereas PT delivers a superior fit for about 75% of the subjects. Moreover, the subjects identified as SEUT types value outcomes linearly and, with only a few exceptions, coincide with the subjects reported to behave risk neutrally by Bruhin et al. (2007), i.e. subjects identified as expected utility types. The participants assigned to the other group seem to distort probabilities by a pattern which is best explained by PT, rather than SEUT.

Thus, previous results on individual heterogeneity that, on average, about one forth of the population seems to behave almost risk neutrally, whereas the majority shows a pronounced fourfold pattern in risk taking behavior are confirmed. Furthermore, even when SEUT fits into general microeconomic theory and describes aggregate choices quite well but not without exceptions, the rigid patterns it imposes on deviations from expected utility seems to prevent it from outperforming PT in a finite mixture context. Consequently, as soon as individual heterogeneity is taken into account, SEUT neither outperforms PT nor does it deliver any additional qualitative insights.

The paper is structured as follows: Section 4.2 discusses the experimental setup and the procedure applied to elicit the certainty equivalents. Section 4.3 explains the two theories' formulations as econometric models for representative choice before it introduces the finite mixture specification. Some estimation issues typical to finite mixture models are also briefly addressed in this part. In section 4.4 the results of the representative choice models as well as the finite mixture model are interpreted. Finally, 4.5 sums up and concludes.

Table 4.1: Differences in Experimental Design

	Zurich 06	Beijing 05
Number of:		
Subjects	118	153
Lotteries	40	28
Observations	4,669	4,281
Procedure	computerized	paper and pencil

4.2 Experimental Design

The study uses experimental data from Bruhin et al. (2007). The experiments were conducted in Zurich 2006 and in Beijing 2005. The participants for the Zurich experiment were randomly selected from the subject pool of the Institute for Empirical Research in Economics consisting of students from various fields of the University of Zurich and the Swiss Federal Institute of Technology Zurich. The Chinese subjects were recruited by flier among the students of Peking University and Tsinhua University. As both experiments have the same basic design in common, this section presents the Zurich experiment in detail and discusses in which respects the experimental design in Beijing 2005 deviates. Table 4.1 summarizes the most important differences between the two experiments.

The experiments both aimed at eliciting participants' certainty equivalents for 28 to 40 two-outcome lotteries. One half of the lotteries were framed as choices between risky and certain gains ("gain domain"), the other half as options between risky and certain losses ("loss domain"). For each lottery in the loss domain the participants received an initial monetary endowment to cover their potential losses.

The certainty equivalents were elicited by applying the following choice menu (Kahneman et al., 1991): For any given lottery under consideration the decision sheet contained two options, the lottery and a certain outcome which varied in 20 equal steps from the lottery's maximum payoff to the lottery's minimum payoff, as shown in Figure 4.1. For each row the subjects had to reveal whether they prefer the lottery or the actual certain payoff. The certainty equivalent was calculated as the arithmetic mean between the smallest certain amount preferred to the lottery and the subsequent certain amount, were the lottery was first chosen. In the example depicted in Figure 4.1 the subject's choices

Figure 4.1: Design of the Decision Sheet

Decision situation:
22

	Option A	Your Choice:				Option B Guaranteed payoff amounting to:
1		A		o	B	20
2		A		o	B	19
3		A		o	B	18
4		A		o	B	17
5		A		o	B	16
6		A		o	B	15
7	A profit of CHF 20 with	A		o	B	14
8		A	o		B	13
9	probability 75%	A	o		B	12
10		A	o		B	11
11	and a profit of CHF 0 with	A	o		B	10
12		A	o		B	9
13	probability 25%	A	o		B	8
14		A	o		B	7
15		A	o		B	6
16		A	o		B	5
17		A	o		B	4
18		A	o		B	3
19		A	o		B	2
20		A	o		B	1

OK

are indicated by the small circles implying a certainty equivalent of 13.5 Swiss Francs.[1] The experiment conducted in Zurich used a computerized procedure programmed in the software z-Tree (Fischbacher, 2007) whereas in Beijing the decision sheets were printed out on paper. In both experiments the lotteries appeared in random order.

Payoffs per subject averaged out at approximately 31 Swiss Francs and 20 Chinese Yuan, considerably more than a local student assistant's hourly compensation, plus a show up fee of 10 Swiss Francs and 20 Chinese Yuan, respectively, thus generating salient incentives[2]. In Zurich the lotteries' outcomes, x_1 and x_2, varied between zero to 150 Swiss francs. The payoffs in the Beijing 2005 experiment were comparable in terms of typical local stundent's compensation and ranged from zero to 55 Chinese Yuan. Probabilities p of the lotteries' higher gain or loss x_1 varied between 5% and 95%. Table 4.2 shows the two experiments' lotteries in the gain domain.

After reading the instructions, the subjects had to correctly calculate the payoffs for

[1] One Swiss Franc equals about one U.S. dollars.
[2] One Chinese Yuan equals about 0.14 U.S. dollars.

Table 4.2: Gain Lotteries $(x_1, p; x_2)$

Zurich 06						Beijing 05					
p	x_1	x_2	p	x_1	x_2	p	x_1	x_2	p	x_1	x_2
0.05	40	0	0.50	20	0	0.05	15	4	0.75	20	7
0.05	40	10	0.50	20	10	0.05	20	7	0.90	7	4
0.05	50	20	0.50	40	10	0.05	55	20	0.95	15	4
0.05	150	50	0.50	50	20	0.10	7	4	0.95	20	7
0.10	20	10	0.75	40	10	0.25	15	4			
0.10	150	0	0.75	50	20	0.25	20	7			
0.25	40	0	0.90	20	10	0.50	7	4			
0.25	40	10	0.95	40	10	0.50	15	4			
0.25	50	20	0.95	50	0	0.50	20	7			
0.50	10	0	0.95	50	20	0.75	15	4			

Outcomes are denominated in Swiss Francs (Zurich 2006) and Chinese Yuan (Beijing 2005), respectively.

two hypothetical choices before they were permitted to start working on the experimental decisions.[3] In the computerized experiments, there were two trial rounds to familiarize the subjects with the procedure. At the end of the experiment, one row number of one decision sheet was randomly selected for each subject, and the subject's choice in that row determined her payment. The subjects were paid in private afterward. They could work at their own speed, the vast majority of them needed less than an hour to complete the experiment.

4.3 Econometric Models

This section covers the econometric models' formulation and some associated estimation issues. The first two models for fitting aggregate choices are based on a single decision model, SEUT or PT, respectively. The third specification, a finite mixture model, combines these two approaches by simultaneously estimating the behavioral parameters of a group of SEUT as well as PT types. As such a model endogenously determines which of the two decision models best describes a specific subject's choices, it provides an estimate of their respective shares among the population. This procedure yields a basis for deciding whether to take potential heterogeneity into account or to assume a representative

[3] The instructions are available on request.

decision maker.

4.3.1 Prospect Theory for Representative Choice

In PT a subject $i \in \{1, \ldots, N\}$ values any given lottery $\mathcal{G}_g = (x_{1g}, p_g; x_{2g})$, $g \in \{1, \ldots, G\}$, where $|x_{1g}| > |x_{2g}|$, by

$$v\left(\mathcal{G}_g\right) = v\left(x_{1g}\right) w\left(p_g\right) + v\left(x_{2g}\right)\left(1 - w\left(p_g\right)\right). \tag{4.1}$$

The sign-dependent function $v(x)$ denotes how monetary outcomes, x, are valued, whereas $w(p)$ assigns a subjective weight to every outcome probability, p. The gamble's certainty equivalent $\hat{ce}_g$ can be written as

$$\hat{ce}_g = v^{-1}\left[v\left(x_{1g}\right) w\left(p_g\right) + v\left(x_{2g}\right)\left(1 - w\left(p_g\right)\right)\right]. \tag{4.2}$$

To make the model operational both the value function, $v(x)$, and the probability weights, $w(p)$, need to be specified by assuming a functional form. A natural candidate for $v(x)$ is a sign-dependent power function

$$v(x) = \begin{cases} x^{\alpha} & \text{if } x \geq 0 \\ -(-x)^{\beta} & \text{otherwise,} \end{cases} \tag{4.3}$$

which has a convenient interpretation and turned out to be the best compromise between parsimony and goodness of fit in the context of PT (Stott, 2006). The probability weighting curve, $w(p)$, is modeled as two-parameter function as proposed by Goldstein and Einhorn (1987) and Lattimore et al. (1992):

$$w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1-p)^{\gamma}}, \ \delta \geq 0, \gamma \geq 0. \tag{4.4}$$

This specification has not only proven to account well for individual heterogeneity (Wu et al., 2004) but also its parameters have a neat interpretation: The parameter γ largely governs the slope of the curve, whereas the parameter δ largely governs its elevation. The smaller the value of γ, the more strongly the probability weighting function deviates from linear weighting. The larger the value of δ, the more elevated the curve, ceteris paribus. Linear weighting is characterized by $\gamma = \delta = 1$. In a sign-dependent model, the parameters may take on different values for gains and for losses.

As PT explains *deterministic* choice a stochastic error term needs to be assumed in order to estimate the model's parameters based on the elicited certainty equivalents, ce_{ig}. There could be many different sources of error, such as carelessness, hurry or inattentiveness, resulting in wrong answers. Thus, as suggested by Hey and Orme (1994), the model assumes an additive Fechner-type error ϵ_{ig}, such that $ce_{ig} = \hat{ce}_g + \epsilon_{ig}$. The Central Limit Theorem indicates that the errors are normally distributed and simply add white noise. Furthermore, the model has to account for heteroskedasticity in the error variance. For each lottery the subjects have to consider 20 certain outcomes, which are equally spaced throughout the lottery's range $|x_{1g} - x_{2g}|$. Since the observed certainty equivalents ce_{ig} are calculated as the arithmetic mean of the smallest certain amount preferred to the lottery and the subsequent certain amount the measurement error in the model's dependent variable is proportional to the lottery range. This yields the form $\sigma_g = \sigma|x_{1g} - x_{2g}|$ for the standard deviation of the error term distribution, where σ denotes an additional parameter to be estimated.

Given these assumptions on the distribution of the error term, the individual contribution to the model's likelihood can be expressed as

$$f\left(ce_i, \mathcal{G}; \theta\right) = \prod_{g=1}^{G} \frac{1}{\sigma_g} \phi\left(\frac{ce_{ig} - \hat{ce}_g}{\sigma_g}\right), \tag{4.5}$$

where ϕ denotes the density of the standard normal distribution. The vector of parameters, $\theta = (\alpha, \beta, \gamma', \delta', \sigma)'$ is estimated by maximizing the model's likelihood given by the product of (4.5) over all individuals.

4.3.2 Stochastic Expected Utility Theory for Representative Choice

In the standard microeconomic model with deterministic preferences a given lottery $\mathcal{G}_g$ is valued by its expected utility, which implies the following certainty equivalent

$$\hat{ce}_g = u^{-1}\left[u\left(x_{1g}\right)p_g + u\left(x_{2g}\right)\left(1 - p_g\right)\right], \tag{4.6}$$

where $u(x)$ represents a subjective utility function. A convenient parametric specification in terms of interpretability and parsimony is, again, a power function

$$u(x) = \begin{cases} x^{\eta} & \text{if } x \geq 0 \\ -(-x)^{\eta} & \text{otherwise,} \end{cases} , \tag{4.7}$$

where η measures $u(x)$'s curvature.

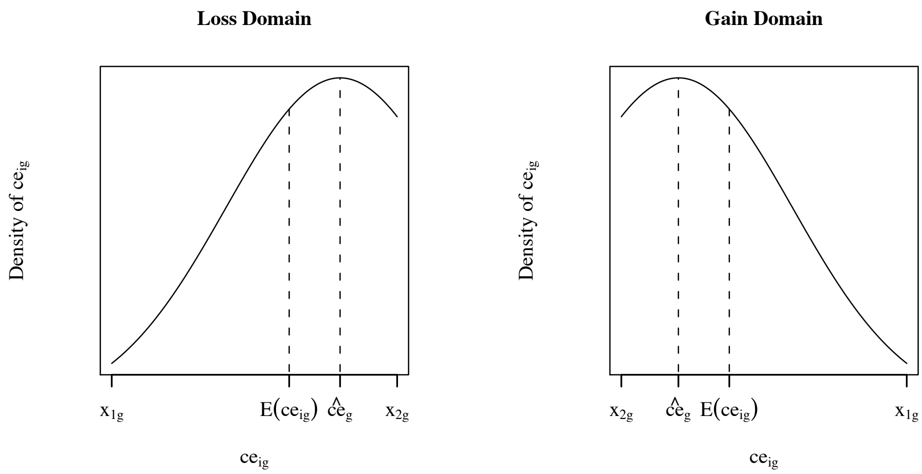
As a purely descriptive theory, SEUT does not aim at explaining the fundamentals underlying such robust phenomena as the fourfold pattern. Hence, subjects do not explicitly distort probabilities, but they are allowed to make random errors when computing the expected utility of a risky lottery. However, as the lottery's most extreme payoffs represent obvious bounds, SEUT assumes that a lottery's value cannot exceed the value of its highest outcome nor can it fall below the value of its lowest outcome. Thus, instead of applying the standard Fechner model with symmetric and unbounded errors, Blavatskyy (2007) suggests truncating the error term, ω_{ig}, at $x_{1g} - \hat{ce}_g$ and at $x_{2g} - \hat{ce}_g$, so that the certainty equivalent $ce_{ig} = \hat{ce}_g + \omega_{ig}$ is limited to lie within the lottery's range, x_{1g} and x_{2g}. As Figure 4.2 illustrates, ce_{ig} can only be symmetrically distributed if $p = 0.5$, and is the more asymmetrically distributed the more p differs from 0.5. Consequently, for $p \neq 0.5$, the expected error $\mathrm{E}(\omega_{ig}) \neq 0$, and ce_{ig} deviates from $\hat{ce}_g$ with a higher probability towards the lottery's center than towards its bounds. So in the gain (loss) domain for $p < 0.5$, the realized certainty equivalent, ce_{ig}, tends to be larger (smaller) than the value predicted by expected utility theory, $\hat{ce}_g$. A decision maker behaving according to SEUT, therefore, still exhibits a specific fourfold pattern in her choices which is driven by stochastic errors, even if she weights probabilities completely linearly.

Analogous to the assumed structure in the PT model, ω_{ig} has a (truncated) normal distribution and is affected by the same source of heteroskedasticity, i.e., its standard deviation is denoted by $\xi_g = \xi \left| x_{1g} - x_{2g} \right|$, with an unknown parameter ξ. Under these assumptions, the individual contribution to the model's likelihood can be written as

$$h\left(ce_i, \mathcal{G}; \psi\right) = \prod_{g=1}^{G} \frac{\frac{1}{\xi_g} \phi\left(\frac{ce_{ig} - \hat{ce}_g}{\xi_g}\right)}{\left| \Phi\left(\frac{x_1 - \hat{ce}_g}{\xi_g}\right) - \Phi\left(\frac{x_2 - \hat{ce}_g}{\xi_g}\right) \right|} , \tag{4.8}$$

where Φ denotes the standard normal's cumulative distribution function. Taking the

Figure 4.2: Distribution of ce_{ig} under SEUT ($p_g = 0.2$, $\xi = 0.4|x_{1g} - x_{2g}|$, $\eta = 1$)



product over all individuals and maximizing the resulting likelihood function yields the maximum likelihood estimates of the model's parameters $\psi = (\eta, \xi)'$.

4.3.3 Finite Mixture Model to Control for Heterogeneity

Since there is evidence for individual heterogeneity in risk taking behavior (Hey and Orme, 1994), aggregating the data and estimating one single decision model veils potentially important behavioral differences and may deliver misleading results. However, estimating all decision models under consideration for each participant separately is highly inefficient and is often rendered impossible by the limited amount of data available per individual. Furthermore, to draw meaningful conclusions, the subjects would still need to be classified by some method into different groups, based on their estimated behavioral parameters.

Thus, instead of operating at the individual level, the finite mixture model proposed here relaxes the assumption of one single representative decision maker by introducing two behavioral types, one PT the other SEUT. *A priori* an individual i's group membership is unknown. Hence, her contribution to the model's likelihood consists of the two decision models' individual likelihoods, (4.5) and (4.8), weighted by the probability that she belongs to the respective type:

$$\ell\left(\Psi; ce_i, \mathcal{G}\right) = \pi_{SEUT}\, h\left(ce_i, \mathcal{G}; \psi\right) + \left(1 - \pi_{SEUT}\right)\; f\left(ce_i, \mathcal{G}; \theta\right)\ , \tag{4.9}$$

where the vector $\Psi = (\eta, \alpha, \beta, \gamma', \delta', \xi, \sigma, \pi_{SEUT})'$ contains all the model's parameters. Note that the probability of being drawn from the SEUT group, π_{SEUT}, equals the fraction of SEUT types among the population and needs to be estimated too. After taking logs, the product over all individuals of (4.9) represents the finite mixture model's log likelihood

$$\ln L\left(\Psi; ce, \mathcal{G}\right) = \sum_{i=1}^{N} \ln\left[\pi_{SEUT}\, h\left(ce_i, \mathcal{G}; \psi\right) + \left(1 - \pi_{SEUT}\right)\; f\left(ce_i, \mathcal{G}; \theta\right)\right]. \tag{4.10}$$

As any finite mixture model's likelihood, (4.10) is highly nonlinear even after taking logs and the log likelihood still contains products, π_{SEUT}, cannot be estimated separately from the two types' average parameters, ψ and θ, respectively. Moreover, (4.10) may be multimodal and unbounded (for more details see McLachlan and Peel (2000) and Render and Walker (1984)), which renders direct maximum likelihood estimation difficult. In order to cope with these problems effectively, the estimation routine, programmed in the R environment (R Development Core Team, 2006), first applies the Expectation Maximization (EM) algorithm (Dempster et al., 1977) before it switches to the much faster BFGS algorithm.[4] The EM algorithm iteratively proceeds in two steps, E and M. During the E-step, an individual *a posteriori* probability of belonging to the SEUT-group, $\tau_{i,SEUT}$, is computed given the actual fit of the data, $\hat{\Psi}$:

$$\tau_{i,SEUT} = \frac{\hat{\pi}_{SEUT}\, h\left(ce_i, \mathcal{G}; \hat{\psi}\right)}{\hat{\pi}_{SEUT}\, h\left(ce_i, \mathcal{G}; \hat{\psi}\right) + \left(1 - \hat{\pi}_{SEUT}\right)\; f\left(ce_i, \mathcal{G}; \hat{\theta}\right)} \tag{4.11}$$

In the following M-step, the model's so called complete data log likelihood is maximized, where $\tau_{i,SEUT}$ replaces unobserved individual group membership. This yields an analytical expression for the relative group size's update, $\hat{\pi}_{SEUT} = 1/N \sum_{i=1}^{N} \tau_{i,SEUT}$, which is computed separately from the model's remaining parameter updates: $\hat{\theta}$ and $\hat{\psi}$. Furthermore, after convergence is achieved, the individual probabilities, $\tau_{i,SEUT}$, obtained at

[4] The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is a Quasi-Newton method which allows solving unconstrained non-linear optimization problems (see for example Broyden (1970)). It is one of the standard hill-climbing optimization routines implemented in the R environment as well as other statistical packages such as STATA.

the maximum likelihood estimate, not only provide a way of endogenously assigning the subjects to either of the two types, but also, they allow to assess how well the two groups are segregated. A clean segregation reflects good performance at capturing individual heterogeneity, whereas relatively high levels of ambiguity in individual group assignment may indicate overfitting, lack of identification, or misspecification.

4.4 Results

The first part of this section discusses the fourfold pattern in risk taking behavior which is found in both data sets. As both theories, PT and SEUT, are able to describe this pattern their goodness of fit for aggregate choices is assessed in a second part. The last part accounts for potential heterogeneity and interprets the finite mixture regressions by inspecting the two type's relative sizes and behavioral parameters, as well as by assesssing the ambiguity in individual group assignment.

The data of both experiments clearly exhibit the fourfold pattern of risk attitudes, which violates expected utility theory. In Figure 4.3, the bars represent the median value of the observed relative risk premia, $RRP = (ev - ce)/|ev|$, where ev denotes the lottery's expected value, sorted by the probability p that the most extreme payoff is realized. Thus, people are risk averse ($RRP > 0$) for small-probability losses and large-probability gains, whereas they are risk seeking ($RRP < 0$) for large-probability losses and small-probability gains.

4.4.1 Representative Choice

For both data sets, Tables 4.3 and 4.4 show the maximum likelihood estimates of the representative PT and SEUT models, respectively. The standard errors, in parentheses, are based on the bootstrap method with 2,000 replications (Efron and Tibshirani, 1993).

In the case of PT, as depicted in Table 4.3, the estimated α and β are both close to unity for the Swiss subjects indicating an almost linear value function. Hence, for gains as well as for losses, the Swiss participants' observed risk attitudes are mostly driven by nonlinear probability weighting: With γ smaller than one and δ close to unity, the

Figure 4.3: Observed Median Relative Risk Premia by p

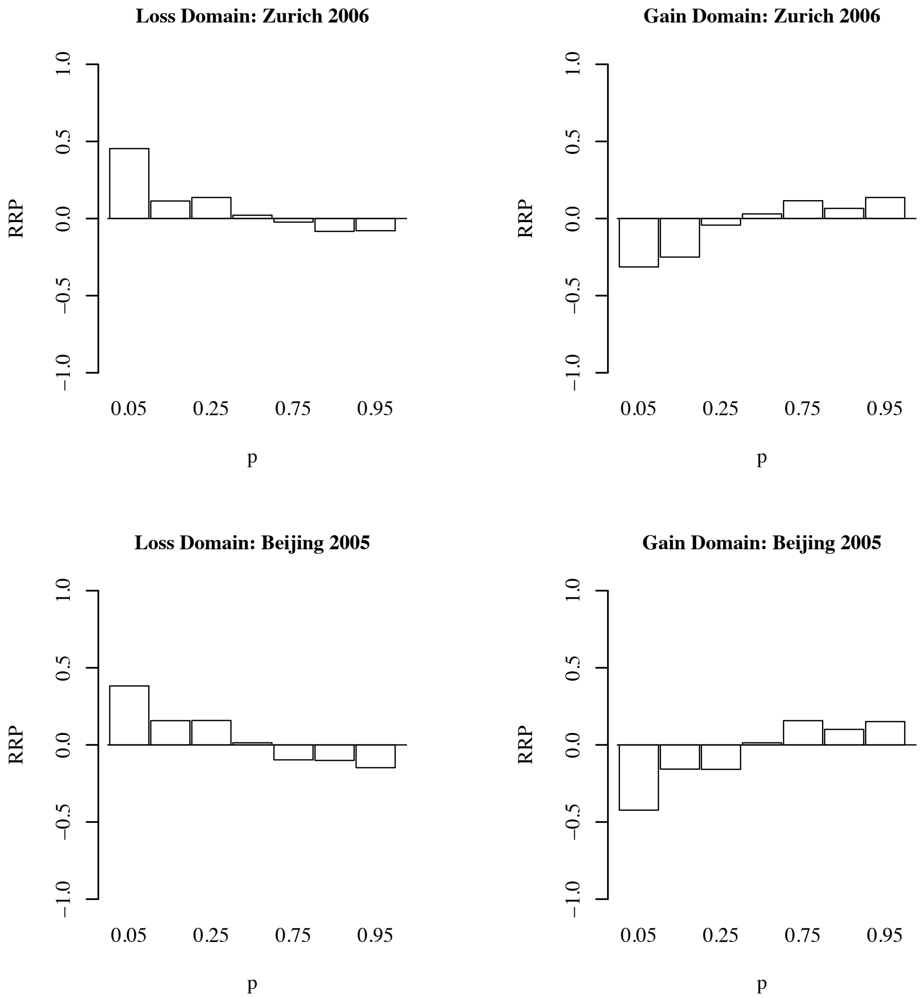


Table 4.3: Prospect Theory Regressions

Parameter Estimates	Zurich 06	Beijing 05
Gain Domain		
α	0.910	0.451
	(0.025)	(0.115)
γ	0.455	0.293
	(0.010)	(0.009)
δ	0.867	1.316
	(0.022)	(0.083)
Loss Domain		
β	1.123	1.202
	(0.045)	(0.127)
γ	0.490	0.352
	(0.010)	(0.009)
δ	1.040	0.887
	(0.037)	(0.163)
σ	0.146	0.163
	(0.002)	(0.002)
$\ln L$	10,089	9,149
BIC	-20,119	-18,239

Standard errors (in parentheses) are based on the bootstrap method with 2,000 replications.

estimated probability weighting curve is inversely S-shaped as typically found in other studies (Stott, 2006; Wu et al., 2004). The Chinese people, on the other hand, value monetary gains at a declining rate of marginal utility ($\alpha = 0.451$), while they weight probabilities much more optimistically than their Swiss colleagues. The smaller value of $\gamma = 0.293$ makes them less sensitive to changes in probabilities, whereas the larger $\delta = 1.316$ corresponds to a generally more elevated probability weighting function. For losses, however, there is no substantial cultural difference, as in both data sets the participants' value functions are only slightly curved and the probability weights are clearly inversely S-shaped. The estimates of σ correspond to an average standard deviation of the error term lying between 14.6% and 16.3% of the lotteries' ranges.

The results for SEUT, on the other hand, are depicted in Table 4.4. The model's only behavioral parameter, η, is estimated to lie in the vicinity of one which implies an almost linear utility function for the Swiss as well as the Chinese subjects. In SEUT, deviations from expected utility theory are directly related to the model's asymmetric error structure. So, the pronounced fourfold pattern observed in the data immediately

Table 4.4: Stochastic Expected Utility Regressions

Parameter Estimate	Zurich 06	Beijing 05
η	0.952	0.947
	(0.016)	(0.059)
ξ	0.211	0.283
	(0.004)	(0.004)
$\ln L$	10,094	8,796
BIC	-20,171	-17,576

Standard errors (in parentheses) are based on the bootstrap method with 2,000 replications.

translates into relatively high estimates for the error's standard deviation, ξ.

Regardless of its more rigid specification, which requires five parameters fewer than PT, SEUT achieves a better fit to the Zurich data even in terms of log likelihood. This is in line with recent findings by Blavatskyy (2007) who reports a good fit of SEUT to several data sets on aggregate choices from different Western countries. In the Beijing data set, however, PT performs better, even when being judged by the Bayesian Information Criterion (BIC), which penalizes its less parsimonious specification ($BIC_{PT} = -18,239$ vs. $BIC_{SEUT} = -17,576$). As mentioned before and indicated by the estimates in Table 4.3, the Chinese subjects judge risky prospects asymmetrically between the gain and loss domain, which cannot be fitted by the model based on SEUT. This may explain SEUT's inferior performance for the Chinese data.

Moreover, since these mixed results on the aggregate level may reflect individual heterogeneity, I assess the models' goodness of fit based on the individual mean squared errors in relative risk premia over all lotteries, $MSE_i = 1/G \sum_{g=1}^{G} \left(\widehat{RRP}_g - RRP_{ig}\right)^2$, which increases in the differences between the predicted $\widehat{RRP}_g$ and the observed RRP_{ig} relative risk premia. Comparing the MSE_i between PT and SEUT indeed reveals some individual heterogeneity in the data: In Zurich and Beijing the share of participants for which SEUT performs better, i.e. delivers smaller MSE_i than PT, amounts to 38% and 36%, respectively. Even though the relative overall performance of SEUT seems to be superior for the Swiss and inferior for the Chinese data, the fraction of people for which it leads to smaller MSE_i is robust. By requiring only two parameters to be estimated SEUT is very parsimonious, but the other side of the coin is that the pattern of devi-

Table 4.5: Finite Mixture Regressions

Parameter Estimates	Zurich 06	Beijing 05
Share of SEUT Types: π_{SEUT}	0.288	0.223
	(0.096)	(0.030)
η	0.974	0.974
	(0.018)	(0.056)
ξ	0.112	0.127
	(0.059)	(0.070)
Share of PT Types: $1 - \pi_{SEUT}$	0.712	0.777
	(0.096)	(0.030)
Gain Domain		
α	0.902	0.377
	(0.035)	(0.132)
γ	0.372	0.212
	(0.027)	(0.013)
δ	0.843	1.371
	(0.033)	(0.099)
Loss Domain		
β	1.167	1.197
	(0.075)	(0.147)
γ	0.398	0.272
	(0.028)	(0.013)
δ	1.029	0.885
	(0.059)	(0.070)
σ	0.151	0.160
	(0.033)	(0.099)
$\ln L$	10,603	9,636
BIC	-21,122	-19,188
ANE	0.048	0.007
Number of Observations	4,669	4,281

Standard errors (in parentheses) are based on the bootstrap method with 2,000 replications.

ations from expected utility theory is rigidly determined by the shape of the truncated normal distribution and its standard deviation (see Figure 4.2 for an illustration). So, its overall performance may react quite sensitively to outliers, domain-specific behavioral asymmetry and the overall composition of the data sets.

4.4.2 Finite Mixture Model

The fact that in both data sets PT and SEUT seem to superiorly fit the subjects' choices by a ratio of about 6:4, suggests using a mix of both theories rather than estimating just one single decision model. Indeed, the BIC reported in Table 4.5 consistently attributes a better performance to the finite mixture regressions than to either of the two representative choice models. Examining the posterior probabilities of individual group membership, $\tau_{i,SEUT}$, allows to assess how well the individual heterogeneity is captured by the assumption of two behavioral types. If the subjects are cleanly segregated all the $\tau_{i,SEUT}$ are either close to zero, indicating members of the PT group, or are close to one, indicating members of the SEUT group. The histograms in Figure 4.4 show the distribution of these probabilities of individual group membership for the two estimated finite mixture regressions. By exhibiting only two prominent spikes, one close to $\tau_{i,SEUT} = 0$ the other close to $\tau_{i,SEUT} = 1$, the histograms reveal graphically that individual group assignment is very clean in both data sets. Another way of assessing the quality of individual group assignment is to look at some measure of entropy which maps the ambiguity in $\tau_{i,SEUT}$ into a single number. For example, the Average Normalized Entropy (El-Gamal and Grether, 1995) defined as

$$ANE = -1/N \sum_{i=1}^{N} \tau_{i,SEUT} \log_2(\tau_{i,SEUT}) + (1 - \tau_{i,SEUT}) \log_2(1 - \tau_{i,SEUT}) \tag{4.12}$$

is normalized to lie within $[0, 1]$. In the case of perfect individual group assignment all $\tau_{i,SEUT}$ equal zero or one, implying $ANE = 0$. $ANE = 1$, on the other hand, reflects complete ambiguity, i.e. all the $\tau_{i,SEUT} = 0.5$, and a failure of classification. Table 4.5 reveals that the ANE only amounts to 0.7% and 4.8% of its maximum value, respectively. These low numbers of entropy, again, reflect the remarkably good performance of the finite mixture model in dealing with individual heterogeneity by cleanly classifying each subject either as a SEUT or a PT type. So, while staying fairly parsimonious the finite mixture model consistently achieves a lower BIC and maps individual heterogeneity very well. Hence for purely statistical reasons, it may be preferred over a representative agent approach.

Whether the individual classification into a SEUT and PT group also bears economic meaning can be assessed on the basis of the corresponding behavioral parameters, ψ and θ, and the mixing proportion, π_{SEUT}. And indeed, by looking at the estimates shown in Table 4.5, a consistent picture emerges:

Figure 4.4: Probability Distribution of Individual Group Membership

The SEUT types are estimated to constitute 28.8% and 22.3% of the population, respectively. Their underlying utility function $u(x)$ is, on average, almost linear with an estimated $\eta = 0.974$. In contrast to the SEUT model for representative choice, the estimates of the error's standard deviation, ξ, amount to only 11.2% and 12.7% of the lotteries' ranges. The deviations from standard expected utility theory are therefore much less pronounced than in the aggregate model. In conjunction with the nearly linear utility function, this implies a behavior much closer to risk neutrality than predicted by the previously discussed SEUT model for representative choice.

To separate the two groups, each subject i in the sample is labeled either as PT type ($\tau_{i,SEUT} < 0.5$) or as SEUT type ($\tau_{i,SEUT} \geq 0.5$) after the estimation of the finite mixture model. After such a separation by type, Figures 4.5 and 4.6 show the median observed relative risk premia sorted by p for the Swiss and Chinese data, respectively. The observed relative risk premia of the participants classified as SEUT types are indeed close to zero over p's entire range, which reflects an almost risk neutral behavior, as illustrated in the lower panels of Figures 4.5 and 4.6. Furthermore, according to Bruhin et al. (2007), the shares of nearly risk neutral participants amount to 22.4% in Zurich 2006 and 20.1% in Beijing 2005 when a mixture model of two different PT types is estimated, instead of assuming one SEUT and one PT type. With the exception of only 8 and 3 subjects,

Figure 4.5: Observed Median Relative Risk Premia by p and Type (Zurich 2006)

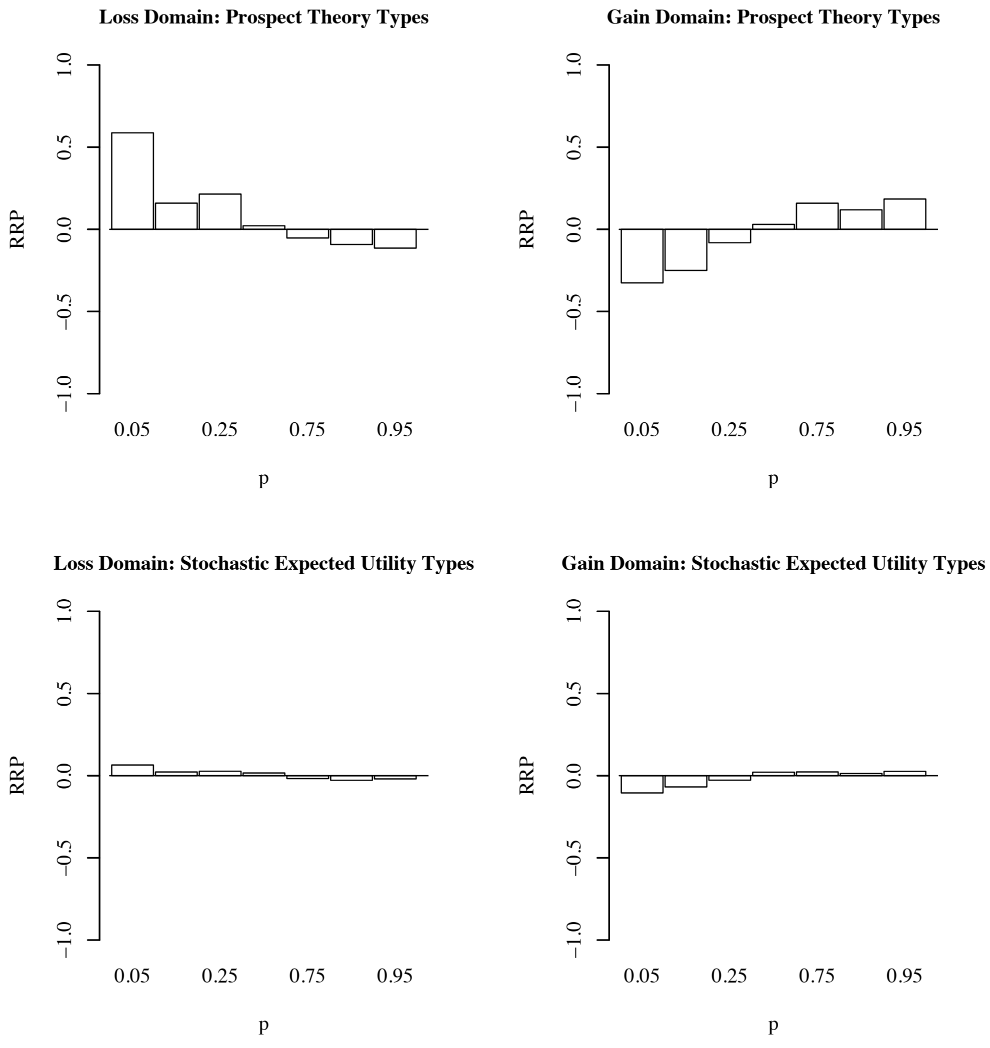
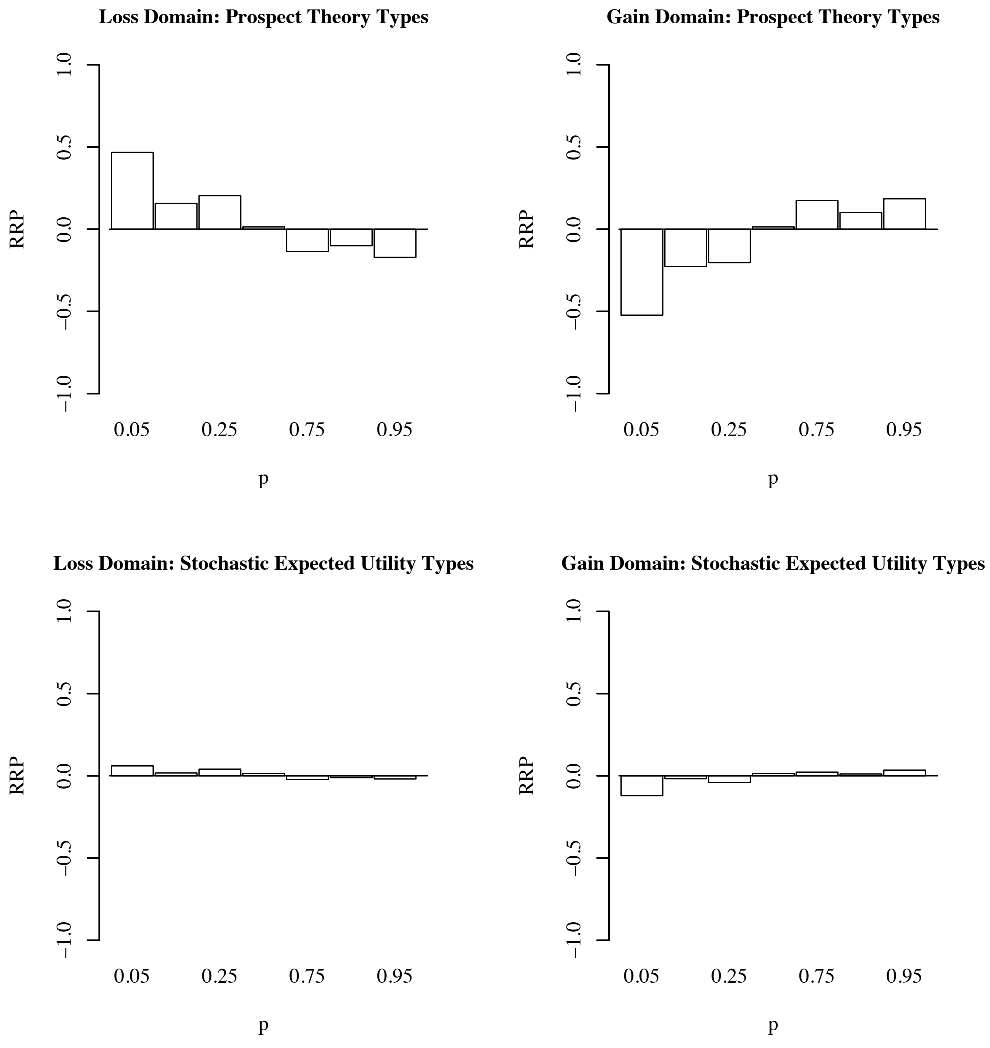


Figure 4.6: Observed Median Relative Risk Premia by p and Type (Beijing 2005)



respectively, the individual classifications found here coincide with the ones reported by Bruhin et al. (2007). So, assuming two behavioral types, one SEUT and the other PT, reproduces previous findings by Conte et al. (2007) and Bruhin et al. (2007) that about one fourth of the individuals can, on average, essentially be characterized as expected value maximizers.

Given the observed relative risk premia as shown in Figure 4.3, it comes at no surprise that the majority of participants, labeled as PT types, exhibit a pronounced fourfold pattern of risk attitudes, as depicted in the upper panels in Figures 4.5 and 4.6. In Zurich as well as Beijing the PT types are estimated to consitute 71.2% and 77.7% of the population, respectively. Their behavioral parameters are qualitatively equivalent to the ones estimated in the PT model for representative choice: The Swiss value functions are only slightly curved ($\alpha = 0.902$, $\beta = 1.167$) whereas, at least for gains, Chinese marginal valuation changes at a steeper rate ($\alpha = 0.377$). Analogous to the case of representative choice, the estimates of γ and δ translate into a probability weighting function exhibiting the characteristic inverse S-shape. And again, when considering small-probability gains, the Chinese participants seem to be more optimistic, as their probability weights are more elevated ($\delta = 1.316$) and less sensitive to changes in p ($\gamma = 0.293$).

Thus, estimating a finite mixture model to account for individual heterogeneity instead of modeling representative choice does not only lead to a better fit to the data but also consistently identifies two distinct behavioral types with a neat economic interpretation: A minority of about 25% behave in an almost risk neutral way, and a majority of about 75% exhibit strong probability distortions best described by a sign-dependent model such as PT. Even though being well suited for describing representative choice, in a finite mixture context SEUT does neither deliver any additional insights nor does it simplify the model's interpretation. In spite of the mixed results for aggregate choice, SEUT is a tempting option if the researcher wants to mildly extend standard microeconomic theory and to parsimoniously model a representative agent's choices. But if markets are imperfect and exhibit strategic complementarity, she wants to avoid aggregation bias and take individual heterogeneity into account (Fehr and Tyran, 2005) and, therefore, may opt for a finite mixture specification where SEUT offers barely any advantages over PT.

4.5 Conclusion

This study compared PT's and SEUT's performance at describing individual decision making under risk in two experimental data sets, one Swiss the other Chinese. On the aggregate level the results are mixed: In conformity with the findings reported by Blavatskyy (2007) for various data sets from other Western countries SEUT clearly outperforms PT in the Swiss data set. In China, however, subjects weight probabilities for gains on average more optimistically and exhibit stronger curvature in their value function. Since SEUT imposes a rigid pattern for deviations from expected utility maximization and cannot cope with behavioral asymmetries between gains and losses, it fits the Chinese data inferiorly when being compared to a more flexible sign-dependent specification such as PT. Furthermore, the finding that both PT and SEUT superiorly describe the choices of a consistent fraction of subjects each calls a representative agent approach into question.

Indeed, the finite mixture regressions, which control for individual heterogeneity by assuming a mix of PT and SEUT types, reveal a coherent picture: In both data sets the mixture model cleanly segregates the subjects into an SEUT and a PT group. Roughly 25% of the individuals are identified as SEUT types and behave essentially risk neutrally, whereas the choices of the remaining 75% are best described by PT. Recent studies estimating finite mixture models based on PT and expected utility theory only report a segregation into two behaviorally similar types (Conte et al., 2007; Bruhin et al., 2007). Moreover, the individual classification found in this study by and large coincides with the one reported in Bruhin et al. (2007). This supports the notion that about one fourth of the subjects can be characterized basically as expected value maximizers.

Despite its parsimony SEUT shows good descriptive power when fitting aggregate choices and, as a modest extension to expected utility theory, molds well into standard microeconomic theory. However, as soon as the often unrealistic assumption of one single representative agent is relaxed, its rigidity causes SEUT to fall short of PT's performance in describing decision making under risk for the majority group of subjects violating expected utility theory. Consequently, its parsimony may make SEUT an elegant option to model aggregate outcomes on perfect markets, but when individual heterogeneity has to be taken into account a flexible sign-dependent specification based on PT is the superior choice.

Bibliography

Bernoulli, D. (1954): "Exposition of a New Theory on the Measurement of Risk," *Econometrica*, 22, 23–36.

Blavatskyy, P. R. (2007): "Stochastic Expected Utility Theory," *Journal of Risk and Uncertainty*, 34, 259–286.

Broyden, C. (1970): "The Convergence of a Class of Double-Rank Minimization Algorithms," *Journal of the Institute of Mathematics and its Applications*, 6, 76–90.

Bruhin, A., H. Fehr-Duda, and T. F. Epper (2007): "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion," *Socioeconomic Institute, University of Zurich, Working Paper*, 0705.

Camerer, C., G. Loewenstein, and D. Prelec (2005): "Neuroeconomics: How Neuroscience can inform Economics," *Journal of Economic Literature*, 43, 6–64.

Camerer, C. F. (1989): "An Experimental Test of Several Generalized Utility Theories," *Journal of Risk and Uncertainty*, 2, 61–104.

Camerer, C. F. and T. Ho (1994): "Violations of the Betweenness Axiom and Nonlinearity in Probability," *Journal of Risk and Uncertainty*, 8, 167–196.

Conte, A., J. D. Hey, and P. G. Moffat (2007): "Mixture Models of Choice Under Risk," *Department of Economics and Related Studies, University of York, Discussion Paper*, 2007/06.

Dempster, A., N. Laird, and D. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

Efron, B. and R. J. Tibshirani (1993): *An Introduction to the Bootstrap*, Chapman and Hall/CRC.

El-Gamal, M. A. and D. M. Grether (1995): "Are People Bayesian? Uncovering Behavioral Strategies," *Journal of the American Statistical Association*, 90, 1137–1145.

Fehr, E. and J. R. Tyran (2005): "Individual Irrationality and Aggregate Outcomes," *Journal of Economic Perspectives*, 19, 43–66.

Fischbacher, U. (2007): "z-Tree: Zurich Toolbox for Readymade Economic Experiments," *Experimental Economics*, 171–178.

Gneezy, U., J. A. List, and G. Wu (2006): "The Uncertainty Effect: When a Risky Prospect Is Valued Less Than Its Worst Possible Outcome," *Quarterly Journal of Economics*, 121, 1283–1309.

Goldstein, W. and H. Einhorn (1987): "Expression Theory and the Preference Reversal Phenomena," *Psychological Review*, 94, 236–254.

Hey, J. D. and C. Orme (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62, 1291–1326.

Houser, D., M. Keane, and K. McCabe (2004): "Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence using a Bayesian Type Classification Algorithm," *Econometrica*, 72, 781–822.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1991): "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias," *Journal of Economic Perspectives*, 5, 193–206.

Kahneman, D. and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–292.

Lattimore, P. K., J. R. Baker, and A. D. Witte (1992): "The Influence of Probability on Risky Choice," *Journal of Economic Behavior and Organization*, 17, 377–400.

McDermott, R., J. H. Fowler, and O. Smirnov (2008): "On the Evolutionary Origin of Prospect Theory Preferences," *Journal of Politics*, forthcoming.

McLachlan, G. and D. Peel (2000): *Finite Mixture Models*, Wiley Series in Probabilities and Statistics.

R Development Core Team (2006): "R: A Language and Environment for Statistical Computing," *R Foundation for Statistical Computing (Vienna, Austria).*

Render, R. A. and X. Walker (1984): "Mixture Densities, Maximum Likelihood and the EM Algorithm," *SIAM Review*, 26, 195–239.

Stahl, D. O. and P. W. Wilson (1995): "On Players' Models of other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10, 218–254.

Starmer, C. (2000): "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk," *Journal of Economic Literature*, 38, 332–382.

Starmer, C. and R. Sudgen (1989): "Probability and Juxtaposition Effects: An Experimental Investigation of the Common Ratio Effect," *Journal of Risk and Uncertainty*, 2, 159–178.

Stott, H. P. (2006): "Cumulative Prospect Theory's Functional Menagerie," *Journal of Risk and Uncertainty*, 32, 101–130.

Trepel, C., C. R. Fox, and R. A. Poldrack (2005): "Prospect Theory on the Brain? Toward a Cognitive Neuroscience of Decision under Risk," *Cognitive Brain Research*, 23, 34–50.

Tversky, A. and D. Kahneman (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

von Neumann, J. and O. Morgenstern (1947): *Theory of Games and Economic Behaviour*, vol. 2nd edn., Priceton, NJ, Priceton University Press.

Wu, G. (1994): "An Empirical Test of Ordinal Independence," *Journal of Risk and Uncertainty*, 9, 39–60.

Wu, G., J. Zhang, and R. Gonzalez (2004): *Decision under Risk*, D. Koehler and N. Harvey (Eds.), The Blackwell Handbook of Judgment and Decision Making, Oxford University Press.

Chapter 5

Happiness Functions with Preference Interdependence and Heterogeneity: The Case of Altruism within the Family

This chapter is joint work with Rainer Winkelmann.

Another version of this chapter has been published as *SOI Working Paper 0702* and is forthcoming in the *Journal of Population Economics*.

5.1 Introduction

Happiness data are increasingly used to tackle important problems in economics, as reviewed by Frey and Stutzer (2002), Layard (2005), or Di Tella and MacCulloch (2006). Indeed, the recent surge in interest is quite dramatic, as pointed out by Clark et al. (2006), who counted 417 happiness-related articles in Econlit between 1960 and 2005, 76% of which had been published since 1995 and 30% since 2003. Most of these papers use, in one way or another, responses to current happiness or life satisfaction questions in cross-section and panel survey data to study the factors motivating individual behavior, as well as the effects of behavior, policies and institutions, on well-being. With the odd exception, much of the previous literature takes a purely individualistic approach to happiness.

The aim of this study is to broaden the existing literature by focusing on positive preference interdependence as in Becker (1981), which may result in altruistic behavior. The question how widely and to what extent altruistic preferences are present in the population is important in various fields of economics. In public economics, the presence of altruism in a substantial fraction of the population may, by adjusting charitable giving and other voluntary transfers, neutralize governmental attempts at redistributing income between generations (Laferrere and Wolff, 2006). In macroeconomic growth modeling, it is crucial to distinguish between two different motivations for intergenerational transfer payments, altruism or joy of giving (Barro, 1974; Bertola et al., 2006). With altruism, individuals' preferences exhibit positive interdependence so that their current utility levels correspond to the discounted utility flows of all future generations, which results in an infinite planning horizon. Individuals motivated solely by joy of giving, however, do not care about the utility of their offspring and, consequently, their bequests will be driven solely by the utility obtained from donating *per se*. This supports an overlapping generations point of view instead of an infinite planning horizon. Moreover, as Fehr and Fischbacher (2002) point out, when markets are imperfect even a minority of people exhibiting some sort of social preferences, such as altruism, can have a major impact on the equilibrium.

In contrast to other studies on altruism, which rely on the analysis of consumption levels and transfers, we focus on subjective well-being as an immediate indicator of utility. Besides being straightforward, this approach allows us to identify altruistic preferences even in a case where the income gap between parents and their children is not wide enough to trigger transfer payments. Imagine a situation where the parents' and their children's marginal utility of income are almost the same. In such a case the parents' marginal disutility of reduced consumption associated with a transfer payment is likely to exceed the marginal utility gained from a transfer induced increase in the children's happiness. So, even if these parents have altruistic preferences in the sense that they care for their children's happiness, this is not revealed in transfer or consumption patterns. However, by directly analyzing the dependence of the parents' utility on their children's subjective well-being our approach allows us to still detect altruistic preferences even in the absence of any transfer payments.

Our analysis is based on data from the German Socio-Economic Panel (GSOEP), a representative annual panel survey initiated in 1984. As the panel population ages, children become adults, move out of their parental homes and set up their own households. The GSOEP has the nice feature that it surveys these descendants' households as well, and thus allows us to generate linked parent-child observations. Between 2000 and 2004, we observe a total of 2,577 interviewed parents with at least one child living in a spin-off household. As these parents are observed in several waves of the panel, and some of them have more than just one adult child who has left home, the number of linkable parent-child pairs amounts to 11,330. Each of these pairs is observed on average for slightly more than 3 years.

Winkelmann (2005), using GSOEP data as well but a different sample including children still at home, reports a long-run correlation of 0.4 to 0.5 in subjective well-being between parents and children. In principle, there are at least three different explanations for this finding: First, attitudes towards well-being may be genetically transmitted. Second, parents and children may share, to some extent, the same environmental and socio-economic attributes. Third, the correlation may be due to a direct, positive, and causal dependence of the parents' utility functions on the utility of their adult children, *i.e.* altruism. In order to isolate the latter effect, we suggest an estimation strategy based on panel data.

We know from experimental economic research that there exist several distinct social preference types, and at least a minority of people seem to exhibit altruistic preferences (see for example Fehr and Schmidt (2002)). Besides reporting vast heterogeneity, Andreoni and Miller (2002), for example, find evidence based on a series of Dictator Games that about 23% of their participants treat their own and the other's payoffs as perfect substitutes, a behavior compatible with altruism. Phelps (2001) conducts Thematic Apperception Tests, a battery of psychological tests aimed at identifying altruistic motivation, and finds that around 20% of the participants responded in an altruistic manner. We will compare our estimates of the prevalence of altruistic preferences, based on survey data, with these findings, gained by applying completely different methodologies in other fields of economic and psychological research.

By estimating a finite mixture regression model we account for unobserved heterogeneity, i.e. the existence of different social preference types, and isolate the share of altruists in a representative household sample. Distinguishing between two preference types allows us to separate the fraction of altruistic parents from the remainder of the sample, which is assumed to behave selfishly.[1] As the finite mixture model endogenously assigns a group membership (altruistic/selfish) probability to each parent, we can test on an individual level how altruism corresponds to transfer payments. This allows us to check the plausibility of the endogenous group assignment, as we expect parents with altruistic preferences to pay - at least on average - higher transfers towards their children.

In Section 5.2 the structure of the data set is discussed in greater detail, and descriptive statistics are provided. Section 5.3 covers the basic econometric model, an extensions to account for household-specific effects, and estimation. Section 5.4 presents and interprets the results, while Section 5.5 concludes.

[1]A related approach has been previously applied by Clark et al. (2005) in the context of estimating the responses of well-being to income changes.

Table 5.1: Data Structure

Number of children living outside the parental household	Number of parent observations 2000	2001	2002	2003	2004	2000-2004
One	1,205	1,279	1,313	1,377	1,458	6,632
Two	341	334	363	373	370	1,781
Three	63	51	61	77	74	326
Four	3	3	7	5	4	22
Five	4	4	2	2	2	14
Total	1,616	1,671	1,746	1,834	1,908	8,775
Total number of parent-child pair observations	2,108	2,132	2,260	2,384	2,446	11,330

5.2 Data Structure and Descriptive Statistics

The analysis is based on the German Socio-Economic Panel (GSOEP), an annual survey of households, which was started in 1984 in West Germany and extended to East Germany in 1990 (Wagner et al., 1993). As mentioned above, it is an important feature of the GSOEP that it follows up on adult children who moved out of their parental homes and may now live in their own families. In more recent waves of the GSOEP, the number of such children living in spin-off households has become large enough to permit empirical studies of parent-child pairs.

We analyze data for the years 2000-2004.[2] In a first step, we extract 2,577 distinct parents with at least one traced child living in a spin-off household. Note that, since for any given parent the number of these children varies between one and five, the number of observed parent-child pairs is higher than the actual number of parents in the data set. Table 1 summarizes the data structure for each wave of the panel. For example, among the 1,616 parents observed in the year 2000 wave, 1,205 parents have only one child living outside the parental household. The remaining 411 parents have several children, so that the total number of observed parent-child pairs adds up to 2,108. The panel is not balanced, as the number of both parents and parent-child pairs varies over time. In total, the data set contains 8,775 parent observations and 11,330 parent-child pair observations.

Beside a broad range of socio-economic variables, the GSOEP provides information

[2]The 2004 wave was the latest release when this research was started. Before 2000, the number of child spin-offs was relatively small, and we therefore took 2000 as our initial year.

on subjective well-being which can be interpreted as a direct measure of individual utility and, thus, is of central interest to this paper. All respondents are asked directly about their general life satisfaction by the following question: "How satisfied are you with your life, all things considered? Please answer according to the following scale: 0 means completely dissatisfied, 10 means completely satisfied". Since general life satisfaction is measured on an ordinal scale, it needs special consideration in regression models, with parent's well-being as dependent variable and children's well-being as explanatory variable. Section 5.3 discusses these issues in greater detail.

For both parents and children, we extract the following characteristics from the data set, which are generally thought of as being important determinants of subjective well-being (see for example Frey and Stutzer (2001)): health, age, employment status, monthly disposable income, household size, marital status, and mean geographical distance between the parental household and the spin-off households. Health is measured on a self-reported five-point scale which is, for simplicity, converted into an indicator variable of good health status for the highest two values. In contrast to other studies, such as Clark and Oswald (1994), who find evidence for an U-shaped effect of ageing on reported subjective well-being with a minimum around 35 years, age is included among the other regressors only in linear form. Since all the parents in the sample are at least 32 years of age, the effect of ageing is expected to be nearly monotonically and positively associated with general life satisfaction. We measure the mean distance in kilometers between the parents' households and their spin-off households based on the geographical coordinate of the country's midpoint, as discussed by Schwarze and Winkelmann (2005). As it is plausible that parents know less about their children the farther away they live, this provides a proxy measure for the parent's general knowledge about their children's living conditions.

Parents may exhibit paternalistic preferences, that is to say, they do not only care about their children's well-being but they may derive direct utility from other attributes of their children, such as education, marital status, and income, regardless of the effect of these attributes on their children's well-being, i.e. for a given level of well-being. If this is the case, adding the children's socio-economic characteristics as controls is crucial for obtaining an unbiased estimator of the prevalence and extent of altruistic preferences.

Table 5.2: Descriptive Statistics

Means (Std.err. in parentheses)	Parents	Children
Subjective well-being[a]	6.573 (0.019)	7.055 (0.020)
Good health (yes=1/no=0)	0.311 (0.005)	0.697 (0.006)
Age	57.4 (0.093)	30.7 (0.075)
Unemployed (yes=1/no=0)	0.085 (0.003)	0.070 (0.003)
Monthly income (in EUR)	4,567 (32.78)	4,030 (25.97)
Female (yes=1/no=0)	0.542 (0.005)	0.513 (0.006)
Years of schooling	11.2 (0.026)	12.3 (0.031)
Household size	2.409 (0.011)	2.470 (0.015)
Married (yes=1/no=0)	0.822 (0.004)	0.460 (0.005)
Transfers paid per year (in EUR)	1,315 (60.14)	
Distance between households (in kilometers)	48.2 (1.137)	
Person-year Observations	8,775[b]	6,606[c]

[a] Measured on an $0, 1, \ldots, 10$ scale.

[b] Excludes multiple person-year observations for parents with several children.

[c] Excludes multiple person-year observations for children with two parents.

Additionally, the data set contains information on the annual amount of monetary transfers paid to the children by their parents. This variable is interesting for two reasons: First, if the parents' motivation for paying transfers is joy of giving or reciprocity instead of altruism, we expect parents' well-being to be positively associated with these transfers ceteris paribus, i.e. for a given level of the child's well-being. Thus, we should include it among the other control variables. Second, after assigning each parent to one of the two groups, it allows us to compare the average transfer payments of the altruistic parents with the selfish ones.

Table 2 reveals that children report, on average, a much better health status than their parents, and the mean difference in age between parents and children is about 27 years. Due to their lower age, but possibly also because of secular trends in cohabitation, fewer children than parents are married. Standard errors are reported in parentheses, and we see that the mean differences are statistically significant.

5.3 Model

5.3.1 Basic Model

Our basic modeling framework is an extension of the standard ordered probit model which allows us to endogenously separate altruistic parents from those who are assumed to be selfish. Let $h_{it} = j$, $j = 0, 1, \ldots, 10$, denote the ordered response of parent i at time t on the 11-point happiness scale. Similarly, v_{it} is the ordered response of parent i's child at time t. If there is more than one child, v_{it} is taken to be the response of parent i's child at time t.

The main object of interest is $P(h_{it} = j|x_{it})$, the conditional probability model for the ordered response of the parents' happiness, where $x_{it} = (x_{it1}, \ldots, x_{itk})'$ is a $(k \times 1)$ vector of determinants of subjective well-being, discussed in the previous section, excluding a constant. If we assume an ordered probit formulation with a linear index function $x_{it}'\beta = x_{it1}\beta_1 + \ldots + x_{itk}\beta_k$, as in McKelvey and Zavoina (1975), we obtain

$$P(h_{it} = j|x_{it}) = \Phi(\kappa_j - x_{it}'\beta) - \Phi(\kappa_{j-1} - x_{it}'\beta)\,, \tag{5.1}$$

where $\kappa_j > \kappa_{j-1}$ are threshold values, and Φ denotes the cumulative density function of the standard normal distribution.

In order to account for heterogeneity in the parents' preference types we introduce an indicator variable, a_i, such that $a_i = 0$ if parent i is selfish and does not care for the well-being of her adult child, and $a_i = 1$ if she is altruistic. For altruistic parents, their children's well-being, v_{it}, becomes one of the determinants of their own utility, and we therefore expect its coefficient, η, to be positive. Whereas for selfish parents, the children's well-being has no effect on their own general life satisfaction. This yields the conditional probability model's basic form

$$P(h_{it} = j|x_{it}, v_{it}, a_i) = \Phi(\kappa_j - x_{it}'\beta - a_i\eta v_{it}) - \Phi(\kappa_{j-1} - x_{it}'\beta - a_i\eta v_{it})\,. \tag{5.2}$$

In this formulation, the child's well-being, v_{it}, enters as an explanatory variable. Since we treat the parents' happiness h_{it} as an ordinal variable, we should, by symmetry, make the same assumption on the child's well-being. It is not immediately obvious how this can be done in a regression context. Rather than including indicator variables for each possible response value (in which case we lose the ordering information), we follow Terza (1987) and replace v_{it} by a cardinalization that is compatible with the ordered probit assumption, i.e., an underlying normally distributed latent linear index v_{it}^*. The children's subjective well-being responses are replaced by their conditional expectations

$$\tilde{v}_{it} = E(v_{it}^*|v_{it} = j) = E(v_{it}^*|\mu_{(j-1)} \leq v_{it}^* < \mu_{(j)}) = \frac{\phi(\mu_{(j-1)}) - \phi(\mu_{(j)})}{\Phi(\mu_{(l)}) - \Phi(\mu_{(j-1)})}, \tag{5.3}$$

where $\mu_{(j)}$s denote the quantiles of the standard normal distribution for the sample cumulative relative frequencies of the eleven response categories $j = 0, 1, \ldots, 10$, and ϕ stands for the standard normal density. To test the robustness of our results we modeled children's well-being as an indicator, which takes on the value 1 if $v_{it} > 4$ and zero otherwise, instead of applying Terza's cardinalization. Besides the obvious loss in efficiency our estimates remained largely unaffected.

We include the well-being index of the representative (=average) child for parents with several children in the above expression. Therefore, we implicitly assume that parents weight their children's well-being equally.[3] To simplify the interpretation of the model, $\tilde{v}_{it}$ is centered around zero which ensures that its effect on the parents' happiness is captured solely by η and does not have any influence on the vector of threshold parameters κ_j.

5.3.2 Extensions

So far the model assumes a pooled data structure and does not take advantage of the fact that the panel data set contains up to five observations on each parent over time. The data's panel structure, however, may help to resolve some of the potential endogeneity problems.

[3]Schwarze and Winkelmann (2005) find that their results remain robust when running the analysis on a subset of parents having a single child. Therefore, the assumption of a representative child seems to be justified.

First, if there is unobserved variation in the parents' permanent consumption levels which is correlated with the children's consumption due to some unobserved time-unvarying family-specific effects, α_i, the children's well-being, v_i, is endogenous. Second, imagine a situation where both parents and children share similar attitudes towards their life satisfaction, like being intrinsically happy or unhappy. Such a correlation, for example due to genetic transmission, generates an endogeneity problem as well. By ignoring these potential sources of endogeneity we would attribute the whole correlation between parents' and their children's happiness to altruistic preferences even when, say by genetic inheritance, intrinsically content parents may tend to have happier children. Consequently, we would overestimate the weight of altruistic preferences.

However, the data's panel structure allows us to isolate that part of the correlation between parents' and their children's happiness which is caused by altruistic preferences, as long as the unobserved other causes, i.e. the family-specific effects α_i, remain constant over time. In a linear regression model we would eliminate α_i and obtain a fixed-effects estimator by either taking first differences or applying the within-transformation. Unfortunately, due the ordered probit's nonlinear form neither is possible. A dummy variable approach is ruled out as well, since it consumes too many degrees of freedom and leads to an incidental parameters problem with inconsistent maximum likelihood estimators.

To be able to address time-unvarying unobservable effects in probit formulations all the same, Mundlak (1978) proposed to model the correlation between the unobserved time-constant effects and the regressors directly. In our case, by assuming that the family-specific effects, α_i, are normally distributed conditional on the individual means, $\bar{x}$ and $\bar{\bar{v}}$, such that

$$\alpha_i|\bar{x}_i, \bar{\bar{v}}_i \sim \mathrm{N}(\bar{x}_i'\delta_1 + \delta_2\bar{\bar{v}}_i, \sigma_\alpha^2)\,, \tag{5.4}$$

their long-run correlation with the dependent variable, h_{it}, can be segregated from the effect of altruistic preferences. As the sum of two normal distributions is again normally distributed, we obtain the following conditional probability model which accounts for family-specific effects:

$$P(h_{it}=j|x_i,v_i,a_i)= \tag{5.5}$$

$$\Phi(\kappa_j - x_{it}'\beta - a_i\eta\tilde{v}_{it} - \bar{x}_i'\delta_1 - \delta_2\bar{\tilde{v}}_i) - \Phi(\kappa_{j-1} - x_{it}'\beta - a_i\eta\tilde{v}_{it} - \bar{x}_i'\delta_1 - \delta_2\bar{\tilde{v}}_i)\,,$$

where η measures the causal effect of the children's on their parents' happiness. Note that all parameters are now scaled by an unidentified but constant factor $(1+\sigma_\alpha^2)^{-1/2}$. This scaling can be safely ignored, as it cancels out, as long as we base our inference on standard errors obtained by the bootstrap method, and interpret the parameter estimates either in terms of marginal probability effects or relative sizes.

5.3.3 Estimation of the Model

In order to estimate the model we have to deal with the fact that we cannot directly observe a given parent's preference type, i.e. *a priori* we do not know whether she is selfish or altruistic. In the following, we discuss our estimation strategy which allows us to overcome this kind of incomplete-data problem. We also briefly address some issues which typically arise during the maximum likelihood estimation of a finite mixture model.

The conditional probability model directly translates into the parent's type-specific density, which can be written as

$$f(h_i|x_i,v_i,a_i)=\prod_{t=1}^{T_i} f(h_{it}|x_i,v_i,a_i)\,. \tag{5.6}$$

As we do not observe the indicator a_i directly, parent i's preference type is unknown *a priori*. Therefore, we have to weight her type-specific density by the probability that she belongs to the corresponding type, which equals this type's relative size. This yields the model's log likelihood function

$$\ln L(\Psi;x,v)=\sum_{i=1}^{N}\ln\left[\pi_a f(h_i|x_i,v_i,a_i=1)+(1-\pi_a)\,f(h_i|x_i,v_i,a_i=0)\right], \tag{5.7}$$

where π_a is the share of altruists in the population and $\Psi=(\beta',\delta',\kappa',\eta,\pi_a)'$ denotes a vector containing all the unknown parameters of the model which need to be estimated. As in any finite mixture model (for a general treatise see McLachlan and Peel (2000)), the relative size of the altruists's group, π_a, cannot be estimated separately from the remaining parameters of the conditional probability model. It is well known that this highly nonlinearform and the potential multimodality, the existence of several local maxima, of the log likelihood function affect the speed of the optimization algorithm negatively, or even prohibit locating the global maximum.

However, if individual group membership a_i were observed, Dempster et al. (1977) show that the so-called complete data log likelihood function would take on the much simpler form

$$\ln \tilde{L}(\Psi; x, v, a) = \tag{5.8}$$
$$\sum_{i=1}^{N} a_i \left[\ln\, \pi_a + \ln f(h_i|x_i, v_i, a_i = 1)\right] + (1 - a_i) \left[\ln (1 - \pi_a) + \ln f(h_i|x_i, v_i, a_i = 0)\right].$$

In this case, the estimated share of altruists, $\hat{\pi}_a = 1/N \sum_{i=1}^{N} a_i$, would be given analytically and the maximum likelihood estimates of the remaining parameters could be obtained separately by numerically maximizing the corresponding type-specific densities.

Dempster and Laird's Expectation Maximization (EM) algorithm now proceeds iteratively in two steps, E and M. During the E-step, given the actual fit of the data, an *a posteriori* probability of being an altruist is obtained for each parent according to Bayes' Law by

$$\tau_{a,i} = \frac{\pi_a f(h_i|x_i, v_i, a_i = 1)}{\pi_a f(h_i|x_i, v_i, a_i = 1) + (1 - \pi_a)\, f(h_i|x_i, v_i, a_i = 0)}\,. \tag{5.9}$$

In the M-step, the complete data log likelihood is maximized, where the unobserved indicator a_i is replaced by these *a posteriori* probabilities of belonging to the altruistic group. Note that, beside being able to deal with the nonlinearity of the log likelihood function, the EM-Algorithm also allows us, based on these $\tau_{a,i}$, to endogenously classify each parent as being either altruistic or selfish.

The problems caused by multimodality can be addressed by implementing a stochastic version of the EM algorithm, such as the Simulated Annealing Expectation Maximization (SAEM) algorithm developed by Celeux et al. (1996). In each iteration, it has a positive probability of leaving a once taken path to convergence and starting over in a different region of the log likelihood function. This results in much higher chances of converging to the global maximum but comes at the cost of even higher computational demands than the standard EM algorithm. The estimation routine, which we programmed in the R environment (R Development Core Team, 2006), therefore uses a hybrid form (Render and Walker, 1984) of the SAEM algorithm, which is more reliable in the detection of the global maximum, and the much faster BFGS algorithm.[4]

The lowest five categories of parents' subjective well-being responses are only sparsely populated, with 11.4 percent of all responses overall. For practical reasons, we collapsed those five responses into a single category, ensuring that during the bootstrap estimation of the standard errors, all response categories contain at least one observation in each subsample, a requirement for estimation of the full model, with a sufficiently high probability. Moreover, in a single index model such as ours, combining categories does not affect the estimator's consistency. The only costs are some loss in efficiency and the impossibility of predicting conditional outcome probabilities for the single components (which is not essential for our research question). As several randomly generated start values all led to the same maximum likelihood estimates the model seems to be well identified.

5.4 Results

In this section, we present the results of total four finite mixture regressions. The first part deals with model selection issues and therefore addresses the question of whether we need to control for family-specific effects and alternative motivations for paying transfers, such as paternalistic preferences, joy of giving, and reciprocity. The second part sheds light on our main research question by discussing the prevalence and extent of altruistic preferences. Finally, we investigate whether parents who get assigned to the group of altruists actually pay higher average transfers to their children.

[4]The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is a Quasi-Newton method which allows solving unconstrained non-linear optimization problems (see for example Broyden (1970)). It is one of the standard hill-climbing optimization routines implemented in the R environment as well as other statistical packages such as STATA.

5.4.1 Model Selection

Table 3 shows the maximum likelihood estimates of four different finite mixture ordered probit models, which all discriminate between selfish and altruistic parents by analyzing the direct dependence of parental utility on children's well-being. The standard errors, in parentheses, are based on the bootstrap method and clustered by individuals to control for possible serial correlation. Not shown in the Table are coefficients on four time dummies in each model that capture a potential time trend in happiness as well as the Mundlak parameter estimates $\hat{\delta}_1$ and $\hat{\delta}_2$ in the family-effects models.

Model 1 represents the baseline as it only uses the parents' socio-economic characteristics as controls and makes no use of the data's panel structure. While still assuming the data to be pooled over time, Model 2 includes the children's socio-economic characteristics as well. Thus, it takes into account that parents may not only care about their children's happiness but obtain utility directly from their offsprings's socio-economic status, too. In such a case, we should control for these paternalistic preferences and prefer Model 2 over Model 1 in order to avoid omitted variable bias. As mentioned in Section 5.3.2, there may exist unobserved family-specific effects which result in an endogeneity problem and lead to biased estimators as well. In contrast to their pooled counterparts, 1 and 2, the models 3 and 4 use the data's panel structure to control for such time-unvarying unobserved effects by applying Mundlak's formulation. They therefore take the potential correlation between the regressors and these effects into account. Consequently, they consistently identify parents with altruistic preferences even when correlated family-specific effects are present. Since the family-effects models include the individual means over time of all regressors, $\bar{x}$ and $\bar{\bar{v}}$, we have to exclude the variables age, years of schooling, and gender (but not their means over time) in order to avoid perfect multicollinearity. The number of parameters therefore rises by 8 if we go from Model 1 to Model 3, and by 13 from Model 2 to Model 4 respectively.

A further potential source of bias, not explicitly considered so far, can arise due to simultaneity, if children's utility depends on their parents' utility as well. To consider the empirical relevance of such a possibility, we performed a Rivers-Vuong-Test (Rivers and Vuong, 1988) in a pooled standard ordered probit model with the children's age and

Table 5.3: Finite Mixture Estimates of Parental Well-being. ($N = 8,775$ observations)

	Pooled over time		Family Effects	
Coefficients and (Std.err.[a])	Model 1	Model 2	Model 3	Model 4
Fraction of altruists $\hat{\pi}_a$	0.277	0.274	0.210	0.214
	(0.026)	(0.032)	(0.033)	(0.033)
Children's well-being in the group of altruists $\hat{\eta}$	0.865**	0.873**	0.918**	0.912**
	(0.064)	(0.068)	(0.094)	(0.083)
Good health	0.762**	0.760**	0.299**	0.300**
	(0.033)	(0.034)	(0.031)	(0.031)
Log-Income	0.480**	0.471**	0.146**	0.140*
	(0.039)	(0.046)	(0.057)	(0.056)
Unemployed	-0.384**	-0.382**	-0.222**	-0.224**
	(0.065)	(0.059)	(0.062)	(0.061)
Married	0.050	0.037	0.244	0.247
	(0.060)	(0.059)	(0.139)	(0.143)
Log-Household size	-0.223**	-0.206**	0.177	0.185
	(0.063)	(0.066)	(0.091)	(0.095)
Distance between households	-0.054**	-0.058**	-0.003	-0.004
	(0.015)	(0.015)	(0.031)	(0.035)
Transfers paid (in 1,000 EUR)	0.005**	0.006*	0.005*	0.005*
	(0.002)	(0.002)	(0.002)	(0.002)
Age[b]	0.185	0.147**		
	(0.023)	(0.036)		
Good health of the average child		0.009		-0.041
		(0.038)		(0.032)
Log-Income of the average child		-0.049		0.032
		(0.041)		(0.045)
Unemployment of the average child		0.013		0.054
		(0.059)		(0.053)
Average child is married		0.036		-0.012
		(0.053)		(0.055)
Log-Household size of the average child		0.016		0.009
		(0.056)		(0.064)
Age of the average child[b]		0.068		
		(0.055)		
Years of schooling of the average child[c]		0.018		
		(0.010)		
Average Child is female[c]		0.019		
		(0.042)		
Log-Likelihood	-14,442.63	-14,434.04	-14,299.88	-14,288.39
BIC	29,067	29,122	28,854	28,949

[a] All standard errors are clustered and obtained on the basis of 300 bootstrap replications.
[b] Only individual means over time are included in models 3 & 4 due to perfect time-dependence.
[c] Only individual means over time are included in model 4 due to time-invariance.
All models additionally contain six threshold parameters and four time dummies.
Models 3 & 4 contain additional parameters for the individual means over time.
Significance codes: **significant at $\alpha = 1\%$; *significant at $\alpha = 5\%$

gender as instruments. The fact that the estimated residuals from the first-stage linear regression of the test were not significant in the ordered probit estimation of the second stage (p-value=0.37), means that the absence of simultaneity bias could not be rejected.

Looking at the estimated parameters in Table 3, we find the results of all four models to be in line with prior findings in the happiness literature. Health and income both show a significant positive effect on parent's subjective well-being, whereas the impact of unemployment is highly significantly negative. As expected, the effect of good health is very large in relative size.[5] With the exception of log-household size, which is insignificant in the family-effects model, all coefficients preserve the same sign. Furthermore, as the parameter estimates in the family effects models only rely on variation within the individuals over time, it comes at no surprise that their standard errors are generally larger than these estimated from the pooled models.

While our main interest is in patterns regarding altruism, to be discussed in detail below, the regressions also provide some evidence for the presence of paternalistic preferences, joy of giving, and reciprocity. A test for the presence of paternalistic preferences comes down to the question whether the coefficients of the children's socio-economic characteristics are jointly significant. Two likelihood ratio tests (model 2 against model 1, and model 4 against model 3) show that the null-hypothesis of the absence of paternalistic preferences has to be rejected (the p-values are 0.028 and 0.042, respectively). Thus we conclude that parents care directly for their children's socio-economic standing, which rules out models 1 and 3. Since the remaining two models, 2 and 4, are not nested, they cannot be tested against each other. A comparison based on the Bayesian Information Criterion (BIC) reveals a slight advantage for the family effects model.

With regards to joy of giving, a necessary condition for such an effect is that transfers increase a parent's happiness *ceteris paribus*, i.e. for a given happiness of the child. This condition is not sufficient, though, as there may be other explanations why transfers can be associated with increased happiness. One is that parents in a better financial situation give more to their children, and they may be happier for that very reason, i.e., the better financial situation, rather than the transfers *per se*. Therefore, it is important to eliminate

[5]The absolute size of the coefficients in the family-effects model cannot be compared directly to their pooled models' correspondents, as they are scaled by an unidentified, but constant factor $(1+\sigma_\alpha^2)^{1/2}$.

this potential confounding effect by controlling for parental income. Second, the observed transfers could be a "pay-back" for received, or anticipated future services that children provide for their parents. We do not observe such services in the data. Hence, we cannot exclude that part of the transfer effect is due to reciprocity rather than joy of giving proper. From model 4 with family effects (p-value=0.028) or model 2 for the pooled panel (p-value=0.015), there is evidence that transfers have a statistically significant positive effect on well-being, after controlling for income as well as the child's utility. Thus, joy of giving and/or reciprocity appear to be motives for transfers as well.

5.4.2 Prevalence and Extent of Altruistic Preferences

The main parameters of interest, $\hat{\pi}_a$, the estimated fraction of altruists, and $\hat{\eta}$, the extent of interdependence in the altruists' preferences, are highly significant with p-values close to zero in all models. The estimated fraction of altruists is larger (27.4%) in the pooled model than in the model which accounts for family-effects (21.4%), although the difference is not statistically significant. All in all, the estimated share of altruists is comparable in magnitude to the 20% reported by Phelps (2001) who relies on psychological tests in a U.S. survey. So, even if we apply a completely different methodology and examine members of the same family instead of strangers, we obtain results that are qualitatively similar to those of Phelps'. Furthermore, our estimates for the spread of altruism are also similar to the fraction of people who treat their own and others' payoffs as perfect substitutes in dictator games (Andreoni and Miller, 2002). This indicates that, after controlling for children's socio-economic characteristics and parents' income as well as applying a family-effects estimator, survey based estimates can provide some meaningful information on preference interdependence and altruism.

As in any other standard ordered probit model, only the signs of the coefficients within a certain group of the finite mixture ordered probit model have a direct interpretation (Boes and Winkelmann, 2006). Arguably, the most intuitive way of interpreting the quantitative effect of the representative child's well-being in the group of altruistic parents is to compute its average marginal probability effect (AMPE) of observing a certain parental response with regard to well-being. To compute the AMPE, each parent has to

Table 5.4: Average Marginal Probability Effect of $\tilde{v}_{it}$

Response category	Model 2		Model 4	
	Estimates	(Std.err.[a])	Estimates	(Std.err.[a])
Zero to four	-0.155	(0.015)	-0.163	(0.017)
Five	-0.052	(0.007)	-0.045	(0.009)
Six	-0.017	(0.005)	-0.012	(0.006)
Seven	0.009	(0.006)	0.011	(0.008)
Eight	0.087	(0.008)	0.078	(0.012)
Nine	0.060	(0.005)	0.055	(0.005)
Ten	0.069	(0.011)	0.075	(0.013)

[a] All standard errors are clustered and obtained on the basis of 300 bootstrap replications.

be classified either as being altruistic or selfish. This is achieved by assigning each parent to the altruists whose *a posteriori* probability, $\tau_{a,i}$, is greater than 50%. By definition marginal probability effects are zero in the group of selfish parents.

Table 4 shows the AMPE of the child's well-being in the group of altruistic parents.[6] For example, a permanent unit increase in $\tilde{v}_{it}$ (i.e. a one standard deviation increase) would, ceteris paribus, boost the probability of observing the most frequent subjective well-being response, $h = 8$, by 8.7 percentage points in model 2 and 7.8 percentage points in model 4.

5.4.3 Transfer Payments by Preference Type

If the model correctly identifies the parents with altruistic preferences, we expect them to be on average more likely to make transfers to their children. Even though, as argued before, not all the parents in the altruistic group necessarily need to pay actual transfers. As we have both the transfer payments and the individual probabilities of being an altruist we can run a regression to check and quantify this association.

Table 5 shows the results of two OLS regressions of the annual transfer amount, paid by the parents to their representative child, on the *a posteriori* probabilities $\tau_{a,i}$ from models 2 and 4. These regressions control for various socio-economic characteristics of the parents

[6] By definition, the AMPEs have to sum up to zero in both models. The small differences (0.001) from zero in the results reported in table 4 are due to rounding error.

Table 5.5: Regressions of Transfer Amount ($N = 8,775$ observations)

	OLS regression of transfers (in 1,000 EUR)	
Coefficients and (Std.err.[a])	Model 2b	Model 4b
A posteriori probability of being an altruists $\tau_{a,i}$	0.905* (0.400)	1.000* (0.490)
Good health	-0.128 (0.135)	-0.122 (0.132)
Log-Income	1.616** (0.178)	1.619** (0.219)
Unemployed	-0.056 (0.139)	-0.053 (0.143)
Married	0690** (0.166)	0.686** (0.160)
Log-Household size	-1.634** (0.213)	-1.636** (0.228)
Years of schooling	0.251** (0.038)	0.251** (0.043)
Good health of the average child	-0.128 (0.176)	-0.124 (0.180)
Log-Income of the average child	-0.509** (0.178)	-0.509** (0.179)
Unemployment of the average child	0.480 (0.532)	0.481 (0.533)
Average child is married	0.690 (0.190)	0.221 (0.195)
Log-Household size of the average child	0.087 (0.170)	0.085 (0.180)
Years of schooling of the average child	0.047 (0.037)	0.047 (0.037)
Intercept	-10.901** (1.933)	-10.883** (1.780)
R^2	0.049	0.049

[a] All standard errors are clustered and obtained on the basis of 300 bootstrap replications.
Significance codes: **significant at $\alpha = 1\%$; *significant at $\alpha = 5\%$

and their children. As expected, parents' income shows a significant positive sign, whereas the average child's income is negatively correlated with transfers paid by the parents. Parental household size also shows the expected negative sign, and parents with higher education seem to be more willing to pay transfers to their children. Most interestingly, the results show a significant positive relationship between transfer payments and the individual *a posteriori* probabilities of having altruistic preferences.[7] In both models, the estimated transfer amount is roughly $1,000$ Euros higher for altruistic parents than it is for the rest of the population. The fact that parents to whom the model assigns a high probability of having altruistic preferences indeed pay, on average, much higher transfers to their children, gives us a strong indication that the econometric model is capable of identifying the altruists in the data set.

5.5 Conclusion

In this paper, we estimate the share of parents with altruistic preferences in a data set stemming from a representative annual survey, the GSOEP. The panel structure of the data allows us to control for various sources of bias, such as paternalistic preferences, genetically transmitted inclinations towards general life satisfaction or any other sort of time-invariant family-specific effects. The estimated share of altruists lies between 21% to 27% of the population, depending on whether the model accounts for family-specific effects or not. When we control for family-specific effects the estimated fraction of altruists, which lies around one fifth, coincides roughly with the findings of two recent studies relying on different, psychological (Phelps, 2001) as well as experimental (Andreoni and Miller, 2002), methodologies and data sets.

The estimated size of the effect of the children's reported life satisfaction on their altruistic parents' subjective well-being is both robust and relatively large in terms of marginal probability effects. Besides altruism, we find evidence that joy of giving and/or reciprocity provide an additional motivation for parents to pay transfers to their children.

Furthermore, we have shown that actual transfers to the children are on average con-

[7] If we exclude transfers in model 2 and 4, the classification and, consequently, the results remain stable. Therefore, the dependence of $\tau_{a,i}$ on transfers paid seems negligible.

siderably larger for parents who get, with a high probability, assigned to the altruistic group. This provides strong evidence that the econometric model, on average, correctly identifies the parents with altruistic preferences as these individuals show a consistent behavior in their transfer payments. Our approach, which is based on a finite mixture model to account for heterogeneity in preference types and relies on subjective well-being as immediate proxy for utility, seems therefore to be well suited to estimate the share of altruists in panel surveys such as the GSOEP.

Finally, the finding that some parents' subjective well-being positively depends on the happiness of their children living in spin-off households confirms the results of other studies that altruistic preferences are present in at least a minority of the population. While this study focuses on altruism, further research has to show whether other cleanly segregated social preference types exist and how they relate to existing theories of other-regarding preferences. Such a deeper understanding of the heterogeneity in preferences may be crucial in determining equilibria, especially when markets are imperfect. So far, we conclude that altruistic preferences are substantial in their prevalence as well as their extent, and they are likely to play an important role in public economics.

Bibliography

Andreoni, J. and J. Miller (2002): "Giving According to GARP: An Experimental Test of The Rationality of Altruism," *Econometrica*, 70, 737–753.

Barro, R. (1974): "Are Government Bonds Net Wealth?" *Journal of Political Economy*, 82, 1095–1117.

Becker, G. (1981): "Altruism in the Family and Selfishness in the Market Place," *Economica*, 48, 1–15.

Bertola, G., R. Foellmi, and J. Zweimueller (2006): *Income Distribution in Macroeconomic Models*, Princeton University Press, Princeton, Oxford.

Boes, S. and R. Winkelmann (2006): "Ordered Response Models," *Allgemeines Statistisches Archiv*, 90, 165–180.

Broyden, C. (1970): "The Convergence of a Class of Double-Rank Minimization Algorithms," *Journal of the Institute of Mathematics and its Applications*, 6, 76–90.

Celeux, G., D. Chauveau, and J. Diebolt (1996): "Stochastic Versions of the EM Algorithm: An Experimental Study in the Mixture Case," *Journal of Statistical Computation and Simulation*, 55, 287–314.

Clark, A., F. Etile, F. Postel-Vinay, C. Senik, and K. V. der Straeten (2005): "Heterogeneity in Reported Well-Being: Evidence from Twelve European Countries," *The Economic Journal*, 115, C118–C132.

Clark, A., P. Frijters, and M. Schields (2006): "Income and Happiness: Evidence, Explanations and Economic Implications," *PSE, Paris, Working Paper*, 24.

Clark, A. and A. Oswald (1994): "Unhappiness and Unemployment," *The Economic Journal*, 104, 648–659.

Dempster, A., N. Laird, and D. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

Di Tella, R. and R. MacCulloch (2006): "Some Uses of Happiness Data in Economics," *Journal of Economic Perspectives*, 20, 25–46.

Fehr, E. and U. Fischbacher (2002): "Why Social Preferences Matter - The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives," *The Economic Journal*, 112, C1–C33.

Fehr, E. and K. Schmidt (2002): "Theories of Fairness and Reciprocity - Evidence and Economic Applications," in *Advances in Economics and Econometrics - 8th World Congress, Economic Society Monographs*, Camebridge University Press, Camebridge.

Frey, B. and A. Stutzer (2001): *Happiness and Economics*, Princeton University Press.

——— (2002): "What can Economists Learn from Happiness Research?" *Journal of Economic Literature*, 40, 402–435.

Laferrere, A. and F. Wolff (2006): "Microeconomic Models of Family Transfers." in *Handbook on the Economics of Giving, Reciprocity and Altruism*, North-Holland, Amsterdam, vol. 2.

Layard, R. (2005): *Happiness: Lessons from a New Science*, Penguin Books, London.

McKelvey, R. and W. Zavoina (1975): "A Statistical Model for the Analysis of Ordinal Level Dependent Variables," *Journal of Mathematical Sociology*, 4, 103–120.

McLachlan, G. and D. Peel (2000): *Finite Mixture Models*, Wiley Series in Probabilities and Statistics.

Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, 69–85.

Phelps, C. (2001): "A Clue to the Paradox of Happiness," *Journal of Economic Behavior and Organization*, 45, 293–300.

R Development Core Team (2006): "R: A Language and Environment for Statistical Computing," *R Foundation for Statistical Computing (Vienna, Austria).*

Render, R. A. and X. Walker (1984): "Mixture Densities, Maximum Likelihood and the EM Algorithm," *SIAM Review*, 26, 195–239.

Rivers, D. and Q. Vuong (1988): "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." *Journal of Econometrics*, 39, 347–366.

Schwarze, J. and R. Winkelmann (2005): "What can Happiness Research tell us about Altruism? Evidence from the German Socio-Economic Panel," *Socioeconomic Institute, University of Zurich, Working Paper*, 0503.

Terza, J. (1987): "Estimating Linear Models with Ordinal Qualitative Regressors," *Journal of Econometrics*, 34, 275–291.

Wagner, G., R. Burkhauser, and F. Behringer (1993): "The English Language Public Use File of the German Socio-Economic Panel," *Journal of Human Resources*, 28, 429–433.

Winkelmann, R. (2005): "Subjective Well-Being and the Family: Results from an Ordered Probit Model with Multiple Random Effects," *Empirical Economics*, 30, 749–761.

Curriculum Vitae

Personal Details

Date of Birth: December 31, 1981
Place of Birth: Zurich, Switzerland
Citizenship: Swiss, Principality of Liechtenstein

Education

Since 06/2006	Doctoral Studies in Economics at the University of Zurich, Switzerland Supervisors: Prof. Rainer Winkelmann, Prof. Ernst Fehr
10/2001–06/2006	Studies in Economics at the University of Zurich, Switzerland
08/1994–01/2001	Kantonsschule Zuercher Oberland, Wetzikon, Switzerland Maturity (Type C: mathematics and natural sciences)

Research Interests

Finite Mixture Regression Models, Individual Risk Taking Behavior, and Other-Regarding Preferences

Publications

2008	Bruhin A., and R. Winkelmann (2008), "Happiness Functions with Preference Interdependence and Heterogeneity: The Case of Altruism within the Family", *Journal of Population Economics*, forthcoming.

Working Papers

03/2008	"Stochastic Expected Utility and Prospect Theory in a Horse Race: A Finite Mixture Approach", SOI Working Paper No. 0803, University of Zurich, http://www.uzh.ch/sts/research/workingpapers/wp0803.pdf
07/2007	"Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size", SOI Working Paper No. 0708, University of Zurich, http://www.uzh.ch/sts/research/workingpapers/wp0708.pdf, (with Helga Fehr, Thomas Epper, and Renate Schubert)
03/2007	"Risk and Rationality: Uncovering Heterogeneity in Probability Distortion", SOI Working Paper No. 0705, University of Zurich, http://www.uzh.ch/sts/research/workingpapers/wp0705.pdf, (with Helga Fehr, and Thomas Epper)
02/2007	"Risk and Rationality: The Effect of Incidental Mood on Probability Weighting", SOI Working Paper No. 0703, University of Zurich, http://www.uzh.ch/sts/research/workingpapers/wp0703.pdf, (with Helga Fehr, Thomas Epper, and Renate Schubert)